Arroyo, M. and Ortiz, M., Local maximum-entropy approximation schemes: a seamless bridge between finite elements and meshfree methods, *International Journal for Numerical Methods in Engineering*, Vol. 65, Issue 13, pp. 2167-2202, 2006

Local *max-ent* approximation schemes: A seamless bridge between finite elements and meshfree methods

M. Arroyo† and M. Ortiz*

Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, CA 91125, USA

SUMMARY

We present a one-parameter family of approximation schemes, which we refer to as *local* max-ent *approximation schemes*, that bridges continuously two important limits: Delaunay triangulation and maximum-entropy statistical inference. Local *max-ent* approximation schemes represent a compromise—in the sense of Pareto optimality—between the competing objectives of unbiased statistical inference from the nodal data and the definition of local shape functions of least width. Local *max-ent* approximation schemes are entirely defined by the node set and the domain of analysis, and the shape functions are positive, interpolate affine functions exactly, and have a weak Kronecker-delta property at the boundary. Local *max-ent* approximation may be regarded as a regularization, or *thermalization*, of Delaunay triangulation which effectively resolves the degenerate cases resulting from the lack or uniqueness of the triangulation. Local *max-ent* approximation schemes can be taken as a convenient basis for the numerical solution of PDEs in the style of meshfree Galerkin methods. In test cases characterized by smooth solutions we find that the accuracy of local *max-ent* approximation schemes is vastly superior to that of finite elements.

KEY WORDS: maximum entropy, information theory, approximation theory, meshfree methods, Delaunay triangulation

1. INTRODUCTION

This paper is concerned with the formulation of approximation schemes that bridge continuously two important limits: Delaunay triangulation and maximum-entropy statistical inference. The resulting basis functions bear similarities with those obtained from the *Moving Least Squares* (MLS) method, which are the basis of a number of meshfree methods for the numerical solution of partial differential equations (e. g. [1], [2], [3]; cf, also, [4] for a review). Despite some similarities, the approach presented in this paper is distinct from MLS methods and presents a number of important advantages over them.

The approximation schemes are entirely defined by the node set and fall into the general class of *convex approximation schemes*. These are schemes based on shape functions that are positive

*Correspondence to: ortiz@aero.caltech.edu
†Current address: marino.arroyo@upc.edu, LaCàN, Universitat Politècnica de Catalunya, C/ Jordi Girona 1-3, Barcelona 08034, Spain.

and interpolate affine functions exactly. An important property of convex approximation schemes is that they have a weak Kronecker-delta property at the boundary. This property greatly facilitates the imposition of essential boundary conditions and can also be exploited in order to glue together domain patches in a fully conforming way. The positivity of the shape functions endows the approximation schemes with useful properties and structure derived from convex geometry, but makes the construction of high order approximants more involved. Extensions beyond first order methods will be pursued in subsequent work. Another avenue for defining higher-order approximation schemes is to combine the present approach with the *partition of unity method* [5, 6].

The specific convex approximation schemes that we investigate represent a compromise—in the sense of Pareto optimality—between two competing objectives:

i) *Unbiased statistical inference* based on the nodal data;
ii) The definition of local shape functions of *least width*.

Objective (i) is classical in information theory and leads to Jaynes' principle of *maximum entropy* [7]. In the present context the least biased shape functions, which we call global *max-ent* approximants, are those that maximize a suitably defined *entropy of the approximation scheme*. By way of contrast, the most local shape functions, in a sense to be made mathematically precise, are found to be affine shape functions supported on a Delaunay triangulation of the node set. Specifically, we define a one-parameter family of smooth convex approximation schemes, which we refer to as *local* max-ent *schemes*, which have global *max-ent* and Delaunay schemes as limiting cases. In particular, *local* max-ent *approximation schemes subsume simplicial finite elements and the Delaunay triangulation as a special case*. Conversely, local *max-ent* approximation may be regarded as a regularization, or—in analogy to statistical mechanics—a *thermalization*, of Delaunay interpolation. The level of thermalization is smoothly controlled by a non-negative parameter that can be a function of position. This spatial dependence enables a seamless transition from meshfree-type approximants to finite elements. An important feature of this regularization is that it effectively resolves the degenerate cases resulting from the lack or uniqueness of Delaunay triangulation. Thus, for node sets for which the Delaunay triangulation is not unique our regularization selects a unique generalized Delaunay approximant in the limit, namely, that which maximizes the approximation entropy.

The local *max-ent* shape functions follow from an unconstrained convex optimization problem at each evaluation point. The size of this problem equals the spatial dimension. In addition, this problem is guaranteed to be solvable on the convex hull of the node set, and its solution is very robust and efficient. Approximants derived from minimization problems have a long tradition and include cubic splines, thin plate splines, MLS approximants and natural neighbor approximants, to name a few.

Local *max-ent* approximation schemes can be taken as a convenient basis for the numerical solution of PDEs in the style of meshfree Galerkin methods (cf, e. g., [4] for a recent review of Galerkin meshfree methods) or, in problems governed by a minimum principle, by constrained, or Rayleigh-Ritz, minimization. We illustrate the performance of local *max-ent* approximation schemes in this type of applications by means of a patch test and two test cases: the standard benchmark problem of a linear elastic built-in cantilever beam loaded at the tip; and the upsetting and extension of a block of compressible neo-Hookean rubber. In both examples we

find that the accuracy of local *max-ent* approximation schemes is vastly superior to that of finite elements, even when the solution cost is carefully factored in.

The structure of the paper is as follows. In § 2 we begin by establishing the properties of general convex approximation schemes, including a weak Kronecker-delta property at the boundary. In § 3 we adopt an information-theoretical viewpoint and introduce the notions of entropy of an approximation scheme and global *max-ent* approximation. The resulting shape functions are of global support and non-interpolating in general. In order to bring these properties under control, in § 4 we introduce the concept of *width* of a shape function. The sum of the widths of the shape functions supplies a measure of the degree of locality of the approximation. We then proceed to introduce the local *max-ent* approximation schemes by recourse to Pareto optimality. A method for the calculation of the shape functions and some properties of the approximation scheme are presented in this section. Applications to the numerical solution of PDEs are presented in § 5. Some concluding remarks are finally collected in § 6.

2. CONVEX APPROXIMATION SCHEMES

All approximation schemes considered in this paper fall within a class that we shall term the class of *convex approximation schemes*. These convex approximation schemes are characterized by the positivity of the shape functions and by being exact on affine functions. These conditions alone do not determine a unique convex approximation scheme, and most of this paper is devoted to the selection of convex approximation schemes that are optimal according to certain ancillary criteria. However, there are a number of desirable properties that are shared by all convex approximation schemes. In this section we proceed to enumerate these properties.

2.1. Approximants as coefficients of convex combinations

Consider a set of distinct nodes $X = \{\boldsymbol{x}_a,\ a = 1, \dots, N\} \subset \mathbb{R}^d$, to be referred to as the node set. Recall [8] that the convex hull of X is the set

$$\mathrm{conv}X = \left\{\boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{x} = \boldsymbol{X}\boldsymbol{\lambda},\ \boldsymbol{\lambda} \in \mathbb{R}^N_+,\ \mathbf{1}\cdot\boldsymbol{\lambda} = 1\right\} \tag{1}$$

where $\mathbb{R}^N_+$ is the non-negative orthant, $\mathbf{1}$ denotes the vector of $\mathbb{R}^N$ whose entries are one, and $\boldsymbol{X}$ is the $d \times N$ matrix whose columns are the coordinates of the position vectors of the nodes in the node set X. Since X is finite, it follows that $\mathrm{conv}X$ is a compact convex polyhedron, or polytope. Let $u : \mathrm{conv}X \to \mathbb{R}$ be a function whose values $\{u_a;\ a = 1, \dots, N\}$ are known on the node set. We wish to construct approximations to u of the form

$$u^h(\boldsymbol{x}) = \sum_{a=1}^{N} p_a(\boldsymbol{x}) u_a \tag{2}$$

where the functions $p_a : \mathrm{conv}X \to \mathbb{R}$ will be referred to as *shape functions*. A particular choice of shape functions defines an approximation scheme. We shall require the shape functions to

satisfy the zeroth and first-order consistency conditions:

$$\sum_{a=1}^{N} p_a(\boldsymbol{x}) = 1, \quad \forall \boldsymbol{x} \in \mathrm{conv}X, \tag{3a}$$

$$\sum_{a=1}^{N} p_a(\boldsymbol{x})\ \boldsymbol{x}_a = \boldsymbol{x}, \quad \forall \boldsymbol{x} \in \mathrm{conv}X. \tag{3b}$$

These conditions guarantee that affine functions are exactly reproduced by the approximation scheme. We note that if $N = d + 1$ and the point set is affinely independent, the consistency conditions uniquely determine the shape functions over the corresponding $d-$simplex. By way of contrast, the shape functions are not uniquely determined by the consistency conditions in general when $N > d + 1$. In addition, we shall require the shape functions be non-negative, i. e.,

$$p_a(\boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{x} \in \mathrm{conv}X,\ a = 1, \ldots, N. \tag{4}$$

The positivity of the shape functions, together with the partition of unity property, allow us to interpret the shape functions as the coefficients of convex combinations. This viewpoint is common in geometric modelling, e.g., in Bézier and B-Spline techniques [9]. Positive linearly consistent approximants have long been studied in the literature [10]. Recent examples include the Natural Element Method shape functions [11] and subdivision schemes [12]. These methods often present a number of attractive features, such as the related properties of monotonicity, the *variation diminishing property* (the approximation is not more "wiggly" than the data), or smoothness preservation [13], of particular interest in the presence of shocks. Furthermore, they lead to well behaved mass matrices. The positivity restriction is natural in problems where a maximum principle is in force, such as in the heat conduction problem. In the present context, the non-negativity requirement is introduced primarily to enable the interpretation of shape functions as probability distributions. It follows from (3a), (3b) and (4) that the shape functions at $\boldsymbol{x} \in \mathrm{conv}X$ define a convex combination of vertices which evaluates to $\boldsymbol{x}$. In view of this property we shall refer to non-negative and first-order consistent approximation schemes as *convex approximation schemes*.

Let $\boldsymbol{p}(\boldsymbol{x})$ denote the vector of $\mathbb{R}^N$ whose components are $\{p_1(\boldsymbol{x}), \ldots, p_N(\boldsymbol{x})\}$. Then, by virtue of the consistency and non-negativity constraints the domain of $\boldsymbol{p}(\boldsymbol{x})$, or feasible set, is

$$\mathcal{P}_{\boldsymbol{x}}(X) = \left\{ \boldsymbol{p} \in \mathbb{R}^N_+ |\ \boldsymbol{X}\boldsymbol{p} = \boldsymbol{x},\ \mathbf{1} \cdot \boldsymbol{p} = 1 \right\}, \tag{5}$$

Evidently, this set is convex. A first question of interest is whether $\mathcal{P}_{\boldsymbol{x}}(X)$ is non-empty, i. e., whether there exist shape functions consistent with the constraints. The following proposition follows directly by comparison of (1) and (5).

Proposition 2.1. *The feasible set $\mathcal{P}_{\boldsymbol{x}}(X)$ is non-empty if and only if $\boldsymbol{x} \in \mathrm{conv}X$.*

It follows from the preceding observations that non-negative and linearly consistent approximation schemes can only be defined on $\mathrm{conv}X$. If the node set is large enough, Carathéodory's theorem states that at least $N - d - 1$ points in X are not necessary in order to express $\boldsymbol{x} \in \mathrm{conv}X$ as a convex combination of points in X. Thus, as expected, convex approximation schemes are not uniquely determined by the node set in general. It is possible to consider domains Ω which are subsets of $\mathrm{conv}X$. However, for simplicity in the present work we will assume that $\Omega = \mathrm{conv}X$ throughout.

2.2. Behavior at the boundary

In interpolating schemes such as Lagrangian finite elements the shape functions satisfy the so-called Kronecker-delta property, i. e., $p_a(\boldsymbol{x}_b) = \delta_{ab}$. This property is particularly useful when solving partial-differential equations numerically, since it renders the imposition of essential boundary conditions straightforward. Most meshfree methods, in particular those based on the MLS approximation, lack the Kronecker-delta property, and, consequently, the approximation on the boundary of the domain may depend on the nodal data of interior nodes. These methods experience difficulty in enforcing essential boundary conditions (cf, e. g., [14]). In this section we study the behavior of general convex approximation schemes at the relative boundary of $\mathrm{conv}X$, $\mathrm{rbd}(\mathrm{conv}X)$, i. e., the boundary of $\mathrm{conv}X$ regarded as a subset of its affine hull. The relative boundary of $\mathrm{conv}X$ coincides with the boundary of $\mathrm{conv}X$ when $\mathrm{aff}(\mathrm{conv}X) = \mathbb{R}^d$. Here aff denotes the affine hull. In particular, we show that all convex approximation schemes possess a weak Kronecker-delta property at the boundary. This Kronecker-delta property greatly facilitates the imposition of essential boundary conditions, which confers convex approximation schemes a distinct advantage over MLS and other meshfree approximation schemes.

We begin by reviewing a few elementary facts concerning the boundary of polytopes. The faces of the polytope $P = \mathrm{conv}X$ can be characterized as the intersections of P with its supporting hyperplanes, in addition to P itself and $\emptyset$, and are themselves polytopes. An equivalent definition of a face of P is a convex subset F of P such that every closed line segment in P with a relative interior point in F has both endpoints (and hence the entire segment) in F [8]. A proper face of P is one that is neither P nor $\emptyset$. The dimension of a face is the dimension of its affine hull. In particular, the 0-dimensional faces of P are called vertices, coincide with its extreme points, and belong to X. We shall denote by $\mathrm{vert}P$ the collection of vertices of the polytope. In addition, $P = \mathrm{conv}(\mathrm{vert}P)$ and, if F is a face of P, it follows that $\mathrm{vert}F = \mathrm{vert}P \cap F$. The relative interiors of the proper faces of P are a partition of $\mathrm{rbd}P$, i. e., they are disjoint and their union is $\mathrm{rbd}P$. The smallest face of P to which $\boldsymbol{x}$ belongs is the contact set of $\boldsymbol{x}$, $C(\boldsymbol{x})$, and is formally defined as the intersection of P with the intersection of all supporting hyperplanes to P at $\boldsymbol{x}$. Its affine dimension is the facial dimension of $\boldsymbol{x}$. The facial dimension of points interior to $\mathrm{conv}X$ is d, while the facial dimension of extreme points is 0. If $\boldsymbol{x} \in \mathrm{rbd}P$, then $C(\boldsymbol{x})$ is a proper face of P.

Proposition 2.2. *Let $\boldsymbol{p}(\boldsymbol{x})$ define a convex approximation scheme with node set X. Let F be a face of* $\mathrm{conv}X$ *and $\boldsymbol{x}_a \notin F$. Then $p_a = 0$ on F.*

Proof. Suppose otherwise, i. e., suppose that there is a point $\boldsymbol{x} \in F$ and a convex approximation scheme $\boldsymbol{p}(\boldsymbol{x})$ such that $p_a(\boldsymbol{x}) \neq 0$. Since

$$\boldsymbol{x} = \sum_b p_b(\boldsymbol{x})\boldsymbol{x}_b = \sum_{b\neq a} p_b(\boldsymbol{x})\boldsymbol{x}_b + p_a(\boldsymbol{x})\boldsymbol{x}_a \tag{6}$$

and $\boldsymbol{x} \neq \boldsymbol{x}_a$, it follows that $\sum_{b\neq a} p_b(\boldsymbol{x}) \neq 0$. Consider the closed line segment

$$[0,1] \ni t \longmapsto t\boldsymbol{y} + (1-t)\boldsymbol{x}_a \in \mathrm{conv}X, \tag{7}$$

where

$$\boldsymbol{y} = \frac{1}{\sum_{b\neq a} p_b(\boldsymbol{x})} \sum_{b\neq a} p_b(\boldsymbol{x})\boldsymbol{x}_b. \tag{8}$$

Then $\boldsymbol{x} \in F$ is a relative interior point of the segment, corresponding to $t = 1 - p_a(\boldsymbol{x})$, and hence the entire segment, including $\boldsymbol{x}_a$, must be contained in F, which contradicts the assumption. □

Remarks:

1. If E is the union of an arbitrary collection of faces of $\mathrm{conv}X$ and $\boldsymbol{x}_a \notin E$, then it follows that $p_a = 0$ on E.
2. The shape functions corresponding to nodes that belong to $\mathrm{relint}(\mathrm{conv}X)$ vanish in $\mathrm{rbd}(\mathrm{conv}X)$.
3. When approximating a function as in Eq. (2), the value of u^h at a face F depends only on the nodal values corresponding to nodes in $X \cap F$. Let X and Y be two node sets such that $\mathrm{conv}X \cap \mathrm{conv}Y$ is a face of both $\mathrm{conv}X$ and $\mathrm{conv}Y$. Then, given a method to select convex approximants, the approximation schemes based on X and Y are conforming (conforming patches).
4. Suppose that a function u defined over $\mathrm{conv}X$ is affine on a face F. Then $u = u^h$ over F provided $u_a = u(\boldsymbol{x}_a) \;\; \forall \boldsymbol{x}_a \in F \cap X$ (exact interpolation of affine functions on faces).
5. If $\boldsymbol{x}_a$ is an extreme point or vertex of $\mathrm{conv}X$, then $p_b(\boldsymbol{x}_a) = \delta_{ba}$, and consequently, $u_a = u^h(\boldsymbol{x}_a)$ (interpolation at extreme points).
6. Let $\boldsymbol{x} \in \mathrm{rbd}(\mathrm{conv}X)$ with contact set $C(\boldsymbol{x})$. If $\boldsymbol{x}_a \notin C(\boldsymbol{x})$, then $p_a(\boldsymbol{x}) = 0$. Thus, choosing a convex approximation scheme in $\mathcal{P}_{\boldsymbol{x}}(X)$ is equivalent to choosing a convex approximation scheme in $\mathcal{P}_{\boldsymbol{x}}(X \cap C(\boldsymbol{x}))$. Note that the latter problem can be formulated in $\mathrm{aff}C(\boldsymbol{x}) - \boldsymbol{x}$, the subspace of $\mathbb{R}^d$ parallel to $C(\boldsymbol{x})$, whose dimension is the facial dimension of $\boldsymbol{x}$, and involves a reduced node set (reduced face problem).
7. If a $n-$dimensional face contains exactly $n+1$ nodes, then the shape functions on that face are the affine shape functions of the simplex defined by those nodes.

Some of these observations are known in different contexts. For instance, the fact that Bézier curves pass through the end control points is a direct consequence of Proposition 2.2.

2.3. Higher-order consistency

A seemingly natural extension of the convex approximation schemes described in the foregoing would be to impose second and higher-order consistency conditions on the shape functions. However, these extensions are not straightforward. In order to demonstrate the source of the difficulty we may simply consider the one-dimensional case. The second-order consistency condition then takes the form

$$\sum_{a=1}^{N} p_a(x) x_a^2 = x^2. \tag{9}$$

Defining an extended point set $Y = \big\{(x_a, x_a^2); a = 1, \ldots, N\big\} \subset \mathbb{R}^2$, it follows that finding non-negative and second-order consistent approximation schemes amounts to defining a convex approximation scheme on the set $\mathcal{P}_{(x,x^2)}(Y)$. We have seen that this set is nonempty iff (x, x^2) belongs to the set $\mathrm{conv}\big\{(x_a, x_a^2); a = 1, \ldots, N\big\}$. In the context of the classical problem of moments, namely, the problem of finding a probability distribution given its first moments [15, 16], that set is known as the moment space. However, due to the strict convexity of the function $f(x) = x^2$, the condition that (x, x^2) be in the set $\mathrm{conv}\big\{(x_a, x_a^2); a = 1, \ldots, N\big\}$

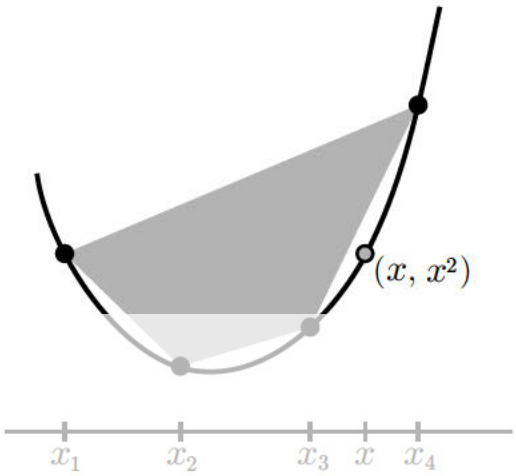


Figure 1. Illustration of the 2nd order moment space in 1D.

cannot be satisfied in general, as illustrated in Fig. 1. Consequently, non-negative convex approximation schemes cannot satisfy Eq. (9). A similar argument applies to higher-order consistency conditions and higher spatial dimensions. This observation notwithstanding, it is nevertheless possible to extend the methods presented here to construct high-order convex approximants, as will be detailed in forthcoming work.

3. GLOBAL *MAX-ENT* APPROXIMANTS

In this section we begin by adopting an information-theoretical viewpoint that naturally leads to a canonical choice of convex approximants, namely, those that maximize the entropy of the approximation scheme. In this framework, the problem of approximating a function from nodal data is regarded strictly as a problem of statistical inference, with no regard given to the physical nature of the data or the mathematical character of the governing field equations. From a strict information-theoretical viewpoint, the overriding concern is to ensure an *unbiased* inference of the function from the data, i. e., one that is free of systematic errors or artifacts.

3.1. Entropy of a convex approximation scheme

Though the principle of maximum entropy is well-known in information theory and statistical physics, in the present context it may stand a brief review. Consider a random variable which can take values in a set of events $\{A_1, A_2, \ldots, A_n\}$ with probabilities $\{p_1, p_2, \ldots, p_n\}$. The set of events and the associated probabilities

$$A = \begin{pmatrix} A_1 & A_2 & \ldots & A_N \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}.$$

are jointly called a finite scheme. We now introduce the concept of entropy—uncertainty—of a given finite scheme, following the introductory text by Khinchin [17]. Consider two finite schemes

$$\begin{pmatrix} A_1 & A_2 \\ 0.5 & 0.5 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} A_1 & A_2 \\ 0.99 & 0.01 \end{pmatrix}.$$

Evidently, the first scheme carries more uncertainty than the second, for which the outcome is almost certainly A_1. The uncertainty associated with a finite scheme can also be interpreted as the amount of information gained by realizing the random variable, thus eliminating completely the uncertainty. Shannon [18] introduced the following measure of uncertainty, or information entropy,

$$H(A) = H(p_1, \dots, p_n) = -\sum_{a=1}^{n} p_a \log p_a \tag{10}$$

with the extension by continuity: $0 \log 0 = 0$. The function $H(A)$ is non-negative, symmetric, continuous, and strictly concave, and possesses a number of properties that are expected of a measure of uncertainty. In particular, $H(\boldsymbol{p}) = 0$ iff one of the probabilities is one and all the others are zero, and attains its maximum for the probabilities $\{1/n, \dots, 1/n\}$, which may intuitively be regarded as the most uncertain or random distribution. Furthermore, $H(1/n, \dots, 1/n) = \log n$, which is an increasing function of n. Consequently, adding events adds uncertainty to this most uncertain distribution. However, adding an impossible event does not alter the level of uncertainty, i. e., $H(p_1, \dots, p_n, 0) = H(p_1, \dots, p_n)$. Suppose that we are given two finite schemes,

$$A = \begin{pmatrix} A_1 & A_2 & \dots & A_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} B_1 & B_2 & \dots & B_m \\ q_1 & q_2 & \dots & q_m \end{pmatrix}.$$

The set of events $A_i B_j$, $i = 1, \dots, n$, $j = 1, \dots, m$ defines a new finite scheme, called product scheme AB. If A and B are independent, we have that $H(AB) = H(A) + H(B)$, whereas if the schemes are dependent, the $H(AB) = H(A) + H_A(B) \leq H(A) + H(B)$, where $H_A(B)$ denotes the expectation of $H(B)$ in scheme A (cf [17] for details). The inequality $H_A(B) \leq H(B)$ can be interpreted by saying that the realization of the scheme A can only decrease the uncertainty of another scheme B. The axiomatic basis of Shannon's information entropy is well-established in information theory (cf, e. g., [17]).

Within the framework just outlined, the entropy of a convex approximation scheme may be defined as follows. Let X be a node set with N nodes, let $\boldsymbol{x} \in \text{conv}X$, and let $\boldsymbol{p}(\boldsymbol{x})$ define a convex approximation scheme. Regard the index set $I = \{1, \dots, N\}$ as a complete system of events. Since the approximation scheme is non-negative and the shape functions add to one, we may regard $\{p_1(\boldsymbol{x}), \dots, p_N(\boldsymbol{x})\}$ as the corresponding probabilities and $H(p_1(\boldsymbol{x}), \dots, p_N(\boldsymbol{x}))$ as the entropy of the corresponding finite scheme.

3.2. Least-biased approximation scheme

An information-theoretical approach to approximation theory can be devised as follows. Equation (3b) is regarded as additional information on the discrete probability distribution $\boldsymbol{p}(\boldsymbol{x})$, namely that the statistical expectation or average of the random variable $\mathfrak{X} : I \to \mathbb{R}^d$, which assigns to each index the position vector of the corresponding node $\mathfrak{X}(a) = \boldsymbol{x}_a$, is $\boldsymbol{x}$. Consistent with this constraint, there are in general multiple probability distributions

$\{p_1(\boldsymbol{x}), \ldots, p_N(\boldsymbol{x})\}$. The problem of approximating a function from scattered data may now be regarded as a problem of statistical inference. From this standpoint, Eq. (2) expresses the expected value $u^h(\boldsymbol{x})$ of a random variable $\mathfrak{U} : I \to \mathbb{R}$ defined by $\mathfrak{U}(a) = u_a$ as determined by the probabilities $\{p_1(\boldsymbol{x}), \ldots, p_N(\boldsymbol{x})\}$.

Suppose that we require that this process of inference be *unbiased*, i. e., that it be based solely on the *a priori* knowledge of the function and free of artifacts or hidden assumptions. According to Jaynes' principle of maximum entropy [7], the least biased probability distribution is that which maximizes entropy subject to all known constraints. Thus, Jaynes states that the maximum entropy distribution is "*... uniquely determined as the one which is maximally noncommittal with regard to missing information. in that it agrees with what is known. but expresses maximum uncertainty with respect to all other matters*". Thus, from a purely information -theoretical viewpoint, the optimal, or least biased, convex approximation schemes are solutions of the program:

$$
\begin{aligned}
(ME) \qquad & \text{maximize} && H(\boldsymbol{p}) = -\sum_{a=1}^{N} p_a \log p_a \\
& \text{subject to} && p_a \geq 0, \quad a = 1, \ldots, N \\
& && \sum_{a=1}^{N} p_a = 1 \\
& && \sum_{a=1}^{N} p_a \boldsymbol{x}_a = \boldsymbol{x}
\end{aligned}
$$

It is interesting to note that in the one-dimensional case this problem gives the *max-ent* solution of the classical problem of moments [19]. Since the information entropy function is strictly concave, its domain $\mathbb{R}^N$, the non-negative orthant, and the constraints are affine, (ME) defines a convex optimization problem. The existence and uniqueness of the solution of this program are established by the following proposition.

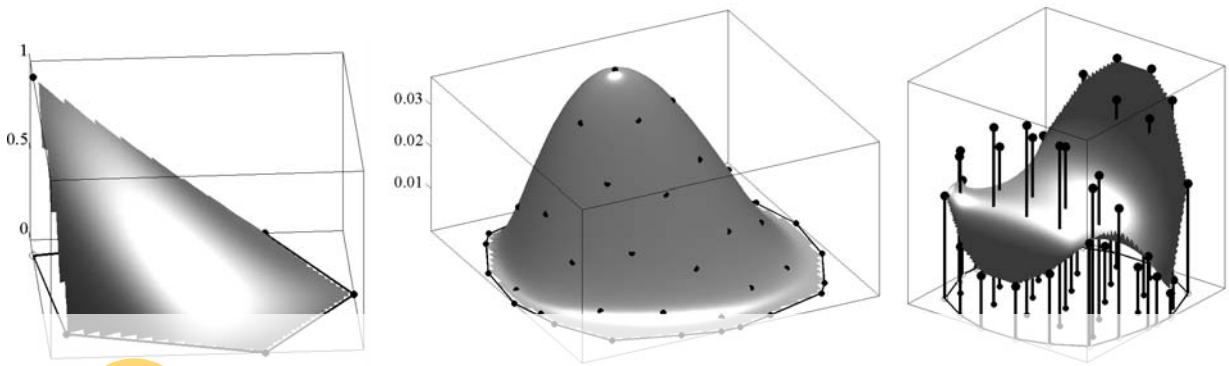
Proposition 3.1. *The program* (ME) *has a solution iff* $\boldsymbol{x} \in \mathrm{conv}X$*, in which case the solution is unique.*

Proof. If $\boldsymbol{x} \in \mathrm{conv}X$, then by proposition 2.1 $\mathcal{P}_{\boldsymbol{x}}(X) \neq \emptyset$. In addition, $\mathcal{P}_{\boldsymbol{x}}(X)$ is a closed and bounded subset of $\mathbb{R}^N$ and, therefore, compact. Hence, by the Weierstrass extreme value theorem $-H$ attains its minimum in $\mathcal{P}_{\boldsymbol{x}}(X)$. Since $-H$ is strictly convex in $\mathcal{P}_{\boldsymbol{x}}(X)$ (the restriction of a strictly convex function to a convex subset) the minimum is unique. □

Since program (ME) depends parametrically on $\boldsymbol{x}$, its unique solution $\boldsymbol{p}_0(\boldsymbol{x})$ is also a function of $\boldsymbol{x}$. We shall refer to the convex approximation scheme defined by $\boldsymbol{p}_0(\boldsymbol{x})$ as the *max-ent approximation scheme*. The smoothness of $\boldsymbol{p}_0(\boldsymbol{x})$ follows as a corollary to proposition 4.2.

3.3. Examples

Given a point set X, the construction of a shape function $p_{0a}(\boldsymbol{x})$ requires solving the problem (ME) for every point $\boldsymbol{x} \in \mathrm{conv}X$. Examples of *max-ent* schemes in the plane are shown in Fig. 2. Figure 2a shows a *max-ent* shape function for a point set consisting of the vertices



(a) (b) (c)

Figure 2. Examples of *max-ent* approximation schemes in the plane. a) Shape function for the vertex of a pentagon; b) shape function for an interior node, illustrating the global character of *max-ent* approximation shemes; and c) *max-ent* approximation, or inference, of a function from scattered data, illustrating the non-interpolating character of *max-ent* approximation schemes.

of a convex pentagon. This example illustrates the *delta Kroneker property* of the *max-ent* shape functions, and the property that the restriction of the *max-ent* shape functions to the edges of the pentagon is linear. Thus, *max-ent* approximation schemes provide a basis for constructing conforming elements in the shape of arbitrary convex polyhedra (cf [20] and [21] for recent alternative methods to construct generalized barycentric coordinates for polyhedra). In recent independent work, maximum entropy methods have been used to construct barycentric coordinates for convex polyhedra, thus defining C^0 approximants on polygonal tesselations [22]. The *max-ent* shape function of an interior node for a larger node set is shown in Fig. 2b. As expected, the shape function vanishes at the boundary. The support of the shape function is highly non-local and extends to the entire convex hull of the node set. In addition, the value of the shape function at its corresponding node differs greatly from unity. Consequently, the *max-ent* approximation is far from interpolating in the interior, and results in a very poor fit to the data as illustrated in Fig. 2c.

This example serves to illustrate some of the limitations of *global max-ent* as a candidate approximation scheme for partial differential equations, namely, its non-local and non-interpolating character. An extension of the *max-ent* concept that provides control over the degree of locality of the shape functions is developed next.

4. LOCAL *MAX-ENT* APPROXIMANTS

As observed in the preceding section, global *max-ent* approximation schemes, while optimal in an information-theoretical sense, are non-local and non-interpolating, which limits their usefulness as approximation schemes for partial differential equations. Control over the degree of locality of *max-ent* approximation schemes can be achieved by adding spatial correlation information in the (ME) program (11). In particular, we wish to control the degree to which

the value of a function at $\boldsymbol{x}$ is correlated to nearby nodal values. Correspondingly, we wish to control the width of the shape functions and their decay with distance away from their corresponding nodes. In this section we extend the *max-ent* framework introduced in the foregoing and build into the approximation scheme these notions of locality.

Define the width of shape function p_a as

$$w[p_a] = \int_\Omega p_a(\boldsymbol{x})|\boldsymbol{x} - \boldsymbol{x}_a|^2 \, d\boldsymbol{x}. \tag{11}$$

where we write $\Omega = \text{conv}X$. Thus, $w[p_a]$ is simply the second moment of p_a about $\boldsymbol{x}_a$. Evidently, other measures of the width of a function can be used instead in order to define alternative approximation schemes. Some alternative measures are briefly discussed in § 4.6. The most local approximation scheme is now that which minimizes the total width

$$W[\boldsymbol{p}] = \sum_{a=1}^N w[p_a] = \int_\Omega \sum_{a=1}^N p_a(\boldsymbol{x}) \; |\boldsymbol{x} - \boldsymbol{x}_a|^2 \, d\boldsymbol{x}, \tag{12}$$

subject to the constraints (3a), (3b) and (4). Since the functional (12) does not involve shape function derivatives its minimization can be performed pointwise. This results in the linear program:

$$\begin{aligned} (RAJ) \quad \text{For fixed } \boldsymbol{x} \text{ minimize} \quad & U(\boldsymbol{x}, \boldsymbol{p}) \equiv \sum_{a=1}^N p_a |\boldsymbol{x} - \boldsymbol{x}_a|^2 \\ \text{subject to} \quad & p_a \geq 0, \;\; a = 1, \ldots, N \\ & \sum_{a=1}^N p_a = 1 \\ & \sum_{a=1}^N p_a \boldsymbol{x}_a = \boldsymbol{x} \end{aligned}$$

An argument identical to that in the proof of Proposition 3.1 shows that the program (RAJ) has solutions if and only if $\boldsymbol{x} \in \text{conv}X$. However, the function $U(\boldsymbol{x}, \cdot)$ is not strictly convex (it is linear) and the solution is not unique in general.

Rajan [23] showed that if the nodes are in general positions (no $(d+1)$ nodes in X are cospherical), then (RAJ) has a unique solution, corresponding to the piecewise affine shape functions supported by the unique Delaunay triangulation associated with the node set X (a Delaunay triangulation verifies that the circumsphere of every simplex contains no point from X in its interior). We shall refer to the convex approximation schemes defined by the solutions $\boldsymbol{p}_\infty(\boldsymbol{x})$ of (RAJ) as *Rajan convex approximation schemes*, and to the approximants corresponding to the piecewise affine shape functions supported by a Delaunay triangulation as *Delaunay convex approximants*. Thus, Rajan's result states that for nodes in general positions, the Delaunay convex approximation scheme coincides with the unique Rajan convex approximation scheme, that is optimal in the sense of the width (11).

When the nodes are not in general positions, the Delaunay triangulation is not unique and the Delaunay approximation schemes are likewise not unique. Since every Delaunay approximation scheme is a Rajan approximation scheme [23], it follows that the latter are likewise not unique. Furthermore, it is readily verified that a convex combination of solutions

of (RAJ) is also a solution. Therefore, the set of solutions $\mathcal{S}_{\boldsymbol{x}}^{RAJ}(X)$ of (RAJ), i. e., the set of Rajan convex interpolation schemes, is a convex subset of $\mathcal{P}_{\boldsymbol{x}}(X)$. Thus, there are Rajan approximation schemes that are not Delaunay approximation schemes and are not associated with a triangulation of the node set. We shall see in Example 4.1 that some Rajan approximation schemes are not even convex combinations of Delaunay approximation schemes. A simple example of non-uniqueness is provided by a node set consisting of four nodes at the corners of a square. Then, the two triangulations corresponding to the two diagonals of the square are Delaunay triangulations and supply solutions of (RAJ). In addition all convex combinations of these solutions are in $\mathcal{S}_{\boldsymbol{x}}^{RAJ}(X)$. This example is further analyzed in Section 4.4. It should be carefully noted that since the approximants are characterized pointwise, their continuity does not follow automatically in the case of non-uniqueness.

4.1. Local max-ent *approximation schemes as a Pareto set*

Thus far we have defined two criteria for selecting convex approximation schemes: maximum entropy and maximum locality, which result in *max-ent* and Delaunay (or Rajan for degenerate node sets) convex approximation schemes, respectively. In general, it is not possible to find convex approximation schemes that maximize both entropy and locality simultaneously, i. e., unbiased estimation and locality are competing objective functions. A standard device for harmonizing such competing objectives is to seek Pareto optima, i. e., convex approximation schemes such that there is none *better*. Specifically, a convex approximation scheme $\boldsymbol{q}$ is better than, or dominates, $\boldsymbol{p}$ iff $-H(\boldsymbol{q}) \leq -H(\boldsymbol{p})$, $U(\boldsymbol{x},\boldsymbol{q}) \leq U(\boldsymbol{x},\boldsymbol{p})$ and at least one of the inequalities is strict. The set of Pareto optima is called the Pareto set. For convex multicriterion optimization, the scalarization of the problem provides a means to find the Pareto set [24]; on the one hand, each solution of the problem:

$$
\begin{aligned}
(LME)_\beta \quad \text{For fixed } \boldsymbol{x} \text{ minimize} \quad & f_\beta(\boldsymbol{x},\boldsymbol{p}) \equiv \beta U(\boldsymbol{x},\boldsymbol{p}) - H(\boldsymbol{p}) \\
\text{subject to} \quad & p_a \geq 0, \;\; a = 1,\dots,N \\
& \sum_{a=1}^{N} p_a = 1 \\
& \sum_{a=1}^{N} p_a \boldsymbol{x}_a = \boldsymbol{x}
\end{aligned}
$$

for $\beta \in (0,+\infty)$ is Pareto optimal. Conversely, each element of the Pareto set is a solution of $(LME)_\beta$ for some $\beta \in (0,+\infty) \cup \{0,+\infty\}$. Note carefully that in general not all the solutions for the values $\{0,+\infty\}$ of β are Pareto optimal. For $\beta = 0$, the uniqueness of the solutions of (ME) guarantees that $\boldsymbol{p}_0$ is Pareto optimal. For $\beta = +\infty$, the linear program (RAJ) is recovered. As already noted, this problem admits multiple solutions in general. From the definition of Pareto optimality, only the element of $\mathcal{S}_{\boldsymbol{x}}^{RAJ}(X)$ with maximum entropy is in the Pareto set. Since $H(\boldsymbol{p})$ is strictly concave in $\mathcal{S}_{\boldsymbol{x}}^{RAJ}(X)$, the Pareto-optimal element for $\beta = +\infty$ is unique and given by:

$$
\boldsymbol{p}_\infty^{Pareto}(\boldsymbol{x}) = \arg \max_{\boldsymbol{p} \in \mathcal{S}_{\boldsymbol{x}}^{RAJ}(X)} H(\boldsymbol{p}). \tag{13}
$$

Since for $\beta \in [0,+\infty)$ the function $f_\beta(\boldsymbol{x},\cdot)$ is continuous and strictly convex in $\mathcal{P}_{\boldsymbol{x}}(X)$, an argument identical to that in the proof of proposition 3.1 shows that $(LME)_\beta$ has a unique

solution $\boldsymbol{p}_\beta(\boldsymbol{x})$ if and only if $\boldsymbol{x} \in \text{conv}X$. We shall refer the convex approximation schemes defined by $\boldsymbol{p}_\beta(\boldsymbol{x})$ as *local max-ent convex approximation schemes*. These convex approximation schemes can be viewed as optimal trade-offs or compromises between information-theoretical optimality and locality.

4.2. Dual problem and exponential form of the shape functions

The problem $(LME)_\beta$ defining the local *max-ent* convex approximations is amenable to analysis by standard duality methods, which also provide a method for the practical calculation of the local *max-ent* approximants. These methods have been extensively applied to *max-ent* problems (cf, e. g., [24]). For simplicity and without loss of generality, throughout this section we shall assume that $\text{aff}X = \mathbb{R}^d$. Under these conditions, the relative interior $\text{relint}(\text{conv}X)$ of $\text{conv}X$ coincides with its interior $\text{int}(\text{conv}X)$. Using the 0th-order consistency condition (3a) it is possible to re-write the 1st-order consistency condition (3b) as

$$\sum_a p_a(\boldsymbol{x}_a - \boldsymbol{x}) \equiv \boldsymbol{Y}\boldsymbol{p} = 0 \tag{14}$$

where $\boldsymbol{Y}$ is the $N \times d$ matrix whose columns are $\boldsymbol{x}_a - \boldsymbol{x}$. The form (14) of the 1st-order condition is preferable to (3b) as it results in better conditioning of the calculations.

Dropping the parametric dependence on $\boldsymbol{x}$ for notational simplicity, the Lagrangian associated with the $(LME)_\beta$ is

$$L(\boldsymbol{p}, \lambda_0, \boldsymbol{\lambda}) = f_\beta(\boldsymbol{p}) + \lambda_0 \left(\mathbf{1} \cdot \boldsymbol{p} - 1\right) + \boldsymbol{\lambda} \cdot \boldsymbol{Y}\boldsymbol{p}, \tag{15}$$

where $\lambda_0 \in \mathbb{R}$ and $\boldsymbol{\lambda} \in \mathbb{R}^d$ are Lagrange multipliers. Therefore, the domain of definition of the Lagrangian is $\mathbb{R}^N_+ \times \mathbb{R} \times \mathbb{R}^d$ and its range is $\mathbb{R}$. Since $\forall \boldsymbol{x} \in \text{conv}X$, $\beta \in [0, \infty)$, $(LME)_\beta$ has a unique solution $\boldsymbol{p}_\beta$, by the Kuhn-Tucker theorem there exist Lagrange multipliers λ_0^* and $\boldsymbol{\lambda}^*$ such that $\left\{\boldsymbol{p}_\beta, \lambda_0^*, \boldsymbol{\lambda}^*\right\}$ is a saddle point of the Lagrangian and verifies the Kuhn-Tucker conditions [8].

In order to analyze the saddle-point problem associated with the Lagrangian (15) it proves convenient to differentiate between interior points, $\boldsymbol{x} \in \text{int}(\text{conv}X)$, and points on the boundary, $\boldsymbol{x} \in \text{bd}(\text{conv}X)$. The interior point case corresponds to Slater's condition that there exist $\boldsymbol{p} \in \mathbb{R}^N_{++}$ such that $\boldsymbol{x} = \boldsymbol{X}\boldsymbol{p}$ (cf Theorem 6.9 in [8]), and, consequently, in this case strong duality holds. We proceed to analyze these two cases in turn.

4.2.1. Interior points Let $\boldsymbol{x} \in \text{int}(\text{conv}X)$ be an interior point. Define the *partition function* $Z : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ associated with the node set X as

$$Z(\boldsymbol{x}, \boldsymbol{\lambda}) \equiv \sum_{a=1}^{N} \exp\left[-\beta|\boldsymbol{x} - \boldsymbol{x}_a|^2 + \boldsymbol{\lambda} \cdot (\boldsymbol{x} - \boldsymbol{x}_a)\right]. \tag{16}$$

Proposition 4.1. *Suppose* $\text{aff}X = \mathbb{R}^d$ *and* $\boldsymbol{x} \in \text{int}(\text{conv}X)$ *and let* $\beta \in [0, \infty)$. *Then, the unique solution of the local* max-ent *problem* $(LME)_\beta$ *is*

$$p_{\beta a}(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x}, \boldsymbol{\lambda}^*(\boldsymbol{x}))} \exp\left[-\beta|\boldsymbol{x} - \boldsymbol{x}_a|^2 + \boldsymbol{\lambda}^*(\boldsymbol{x}) \cdot (\boldsymbol{x} - \boldsymbol{x}_a)\right], \quad a = 1, ..., N, \tag{17}$$

where

$$\boldsymbol{\lambda}^*(\boldsymbol{x}) = \arg\min_{\boldsymbol{\lambda} \in \mathbb{R}^d} \log Z(\boldsymbol{x}, \boldsymbol{\lambda}). \tag{18}$$

Furthermore, the minimizer $\boldsymbol{\lambda}^*(\boldsymbol{x})$ is unique.

Proof. The first N Kuhn-Tucker conditions are

$$0 = \beta|\boldsymbol{x}-\boldsymbol{x}_a|^2 + \log p_{\beta a} + 1 + \lambda_0^* + \boldsymbol{\lambda}^*\cdot(\boldsymbol{x}_a-\boldsymbol{x}), \qquad a = 1,...,N$$

whence we obtain

$$p_{\beta a} = \frac{\exp\left[-\beta|\boldsymbol{x}-\boldsymbol{x}_a|^2 + \boldsymbol{\lambda}^*\cdot(\boldsymbol{x}-\boldsymbol{x}_a)\right]}{\exp(\lambda_0^*+1)},$$

for all $\boldsymbol{x}\in\mathrm{int}(\mathrm{conv}X)$. The optimal Lagrange multipliers λ_0^* and $\boldsymbol{\lambda}^*$ are the maximizers of the Lagrange dual function

$$g(\lambda_0,\boldsymbol{\lambda}) = \inf_{\boldsymbol{p}\in\mathbb{R}^N_+} L(\boldsymbol{p},\lambda_0,\boldsymbol{\lambda}) = -\lambda_0 - f^*_\beta(-\lambda_0\mathbf{1}-\boldsymbol{Y}^T\boldsymbol{\lambda}),$$

where

$$f^*_\beta(\boldsymbol{q}) = \sum_{a=1}^N \exp\left(q_a - \beta|\boldsymbol{x}-\boldsymbol{x}_a|^2 - 1\right).$$

is the function conjugate to $f_\beta(\boldsymbol{p})$ [24]. The Lagrange dual function can be maximized explicitly with respect to λ_0, with the result

$$\lambda_0^* = \arg\max_{\lambda_0\in\mathbb{R}}\left\{-\lambda_0 - \exp(-\lambda_0-1)\sum_{a=1}^N \exp\left[-\beta|\boldsymbol{x}-\boldsymbol{x}_a|^2 + \boldsymbol{\lambda}\cdot(\boldsymbol{x}-\boldsymbol{x}_a)\right]\right\}$$

which in turn yields the identity

$$Z(\boldsymbol{x},\boldsymbol{\lambda}) = \exp(\lambda_0^*+1)$$

Inserting this expression into the Lagrange dual function, the reduced Lagrange dual function

$$\widehat{g}(\boldsymbol{\lambda}) = -\log Z(\boldsymbol{x},\boldsymbol{\lambda})$$

is obtained. Thus it follows that if $\boldsymbol{x}\in\mathrm{int}(\mathrm{conv}X)$, then the local *max-ent* shape functions are given by eqs. (17) and (18). The existence of the minimizer of $-\widehat{g}$, $\boldsymbol{\lambda}^*(\boldsymbol{x})$, is guaranteed by the Kuhn-Tucker theorem. Next we show that this minimizer is unique. The gradient of the objective function is

$$\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda}) \equiv \partial_{\boldsymbol{\lambda}}\log Z(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{a=1}^N p_a(\boldsymbol{x},\boldsymbol{\lambda})(\boldsymbol{x}-\boldsymbol{x}_a), \tag{19}$$

where $p_a(\boldsymbol{x},\boldsymbol{\lambda})$ denotes the evaluation of (17) at an arbitrary value $\boldsymbol{\lambda}$ of the Lagrange multiplier. As expected, the first order optimality condition results in the 1st order consistency condition. The Hessian of the objective function is

$$\boldsymbol{J}(\boldsymbol{x},\boldsymbol{\lambda}) \equiv \partial_{\boldsymbol{\lambda}}\partial_{\boldsymbol{\lambda}}\log Z(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{a=1}^N p_a(\boldsymbol{x},\boldsymbol{\lambda})(\boldsymbol{x}-\boldsymbol{x}_a)\otimes(\boldsymbol{x}-\boldsymbol{x}_a) - \boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda})\otimes\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda}). \tag{20}$$

Consider now a non-zero vector $\boldsymbol{u} \in \mathbb{R}^d$ and, let $u_a = \boldsymbol{u} \cdot (\boldsymbol{x} - \boldsymbol{x}_a)$. Since by assumption $\mathrm{aff} X = \mathbb{R}^d$ it follows that not all u_a are identical. In addition, it follows from (17) that $p_a(\boldsymbol{x}, \boldsymbol{\lambda}) > 0$. Hence, by the strict convexity of the square function, we have

$$\boldsymbol{u} \cdot \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{\lambda}) \boldsymbol{u} = \sum_a p_a(\boldsymbol{x}, \boldsymbol{\lambda})\, u_a^2 - \left(\sum_a p_a(\boldsymbol{x}, \boldsymbol{\lambda})\, u_a \right)^2 > 0.$$

Consequently, $\boldsymbol{J}(\boldsymbol{x}, \boldsymbol{\lambda})$ is positive definite for all $\boldsymbol{\lambda} \in \mathbb{R}^d$ and therefore $\log Z$ is strictly convex and the minimizer of (18) is unique. □

Note that, in the absence of the first order consistency conditions, the resulting approximants are Shepard's functions with Gaussian weight.

4.2.2. Boundary points The treatment of boundary points $\boldsymbol{x} \in \mathrm{bd}(\mathrm{conv} X)$ can be reduced to the problem analyzed in the preceding section by exploiting the *reduced face problem* property of convex approximation schemes (cf § 2.2). Recall that the contact set $C(\boldsymbol{x})$ of the point $\boldsymbol{x}$ with respect to $\mathrm{conv} X$ is the smallest face of $\mathrm{conv} X$ that contains $\boldsymbol{x}$. The dimension of $C(\boldsymbol{x})$ is the face dimension of $\boldsymbol{x}$. The local *max-ent* shape functions then follow from the translation and restriction of problem $(LME)_\beta$ to the node set $X' = X \cap C(\boldsymbol{x}) - \boldsymbol{x}$ and the subspace $L = \mathrm{aff} X'$, whose dimension is the face dimension of $\boldsymbol{x}$. If the face dimension is zero, then the problem is trivial. Otherwise, $\boldsymbol{x} - \boldsymbol{x}$ belongs to the interior of $\mathrm{conv} X' = C(\boldsymbol{x}) - \boldsymbol{x} \subset L$, and proposition 4.1 applies.

4.3. Spacial smoothness of the shape functions

The smoothness of the local *max-ent* approximants $\boldsymbol{p}_\beta(\boldsymbol{x})$ is not guaranteed *a priori*, since they are characterized pointwise by a convex program. We next establish the differentiability of the shape functions with respect to $\boldsymbol{x}$, which ultimately depends on the smoothness of $\boldsymbol{\lambda}^*(\boldsymbol{x})$. As before, we assume $\mathrm{aff} X = \mathbb{R}^d$ for simplicity and without loss of generality.

Proposition 4.2. *Suppose* $\mathrm{aff} X = \mathbb{R}^d$ *and let* $\beta : \mathrm{conv} X \to [0, \infty)$ *be* C^r *in* $\mathrm{int}(\mathrm{conv} X)$. *Then the local* max-ent *shape functions are of class* C^r *in* $\mathrm{int}(\mathrm{conv} X)$.

Proof. Consider the function from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}^d$ given by

$$F(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_a p_a(\boldsymbol{x}, \boldsymbol{\lambda})(\boldsymbol{x} - \boldsymbol{x}_a)$$

If $\beta \in C^r$, then it follows from (16) and (17) that F is likewise is C^r. In the proof of proposition 4.1 we have verified that $F(\boldsymbol{x}, \boldsymbol{\lambda}^*) = 0$ and $\det \partial_{\boldsymbol{\lambda}} F(\boldsymbol{x}, \boldsymbol{\lambda}^*) \neq 0$ for all $\boldsymbol{x} \in \mathrm{int}(\mathrm{conv} X)$. Consequently, by the implicit function theorem, $\boldsymbol{\lambda}^*(\boldsymbol{x})$ is a C^r function in $\mathrm{int}(\mathrm{conv} X)$, and the theorem follows from (16) and (17). □

Computing the derivatives of the shape functions with respect to $\boldsymbol{x}$ is not altogether straightforward because the derivatives of the Lagrange multiplier $\boldsymbol{\lambda}^*(\boldsymbol{x})$ are involved in the calculation. In Appendix II, the explicit form of the derivatives of the shape functions with respect to $\boldsymbol{x}$ is given for a general function $\beta(\boldsymbol{x})$. If β is constant, the remarkably simple expression

$$\nabla p_{\beta a}(\boldsymbol{x}) = -p_{\beta a}(\boldsymbol{x}) \boldsymbol{J}(\boldsymbol{x}, \boldsymbol{\lambda}^*(\boldsymbol{x}))^{-1} (\boldsymbol{x} - \boldsymbol{x}_a). \tag{21}$$

is obtained as a special case.

4.4. Smoothness and limits with respect to the thermalization

We next study the dependence of the local *max-ent* approximants on the thermalization parameter β. We first establish that $\boldsymbol{p}_\beta$ is a continuous function of β in $[0,+\infty)$ and smooth in $(0,+\infty)$. We next analyze the more interesting athermal limit, as $\beta \to +\infty$.

4.4.1. Continuity and differentiability properties

Proposition 4.3. *Let $\boldsymbol{x} \in \mathrm{conv}X$. Let $\boldsymbol{p}_\beta(\boldsymbol{x})$ be the unique minimizer of the problem $(LME)_\beta$. Then, $\boldsymbol{p}_\beta(\boldsymbol{x})$ is a continuous function of β in $[0,+\infty)$.*

Proof. Let $\beta_0 \in [0,+\infty)$. We first show that

$$f_\beta(\boldsymbol{x},\cdot) \longrightarrow f_{\beta_0}(\boldsymbol{x},\cdot) \text{ as } \beta \longrightarrow \beta_0 \text{ uniformly on } \mathcal{P}_{\boldsymbol{x}}(X). \tag{22}$$

We fix $\boldsymbol{x}$ and omit it from all expressions in the proof. Equation (22) simply follows from the fact that for any given $\epsilon > 0$, $|f_{\beta_0}(\boldsymbol{p}) - f_\beta(\boldsymbol{p})| \leq \epsilon, \forall \beta \in [0,+\infty) \cap [\beta_0 - \delta, \beta_0 + \delta]$ with

$$\delta = \epsilon/(N \ \mathrm{diam}^2 X), \tag{23}$$

and $\forall \boldsymbol{p} \in \mathcal{P}(X)$. Here $\mathrm{diam}X$ denotes the diameter of $\mathrm{conv}X$. Then, by uniform convergence and recalling that $(LME)_\beta$ and $(LME)_{\beta_0}$ have unique minimizers, we have

$$\lim_{\beta\to\beta_0} f_\beta(\boldsymbol{p}_\beta) = \lim_{\beta\to\beta_0} \min_{\boldsymbol{p}\in\mathcal{P}(X)} f_\beta(\boldsymbol{p}) = \min_{\boldsymbol{p}\in\mathcal{P}(X)} f_{\beta_0}(\boldsymbol{p}) = f_{\beta_0}(\boldsymbol{p}_{\beta_0}).$$

We now consider a sequence $\{\beta_k\}_{k\in\mathbb{N}} \subset [0,+\infty)$ converging to β_0, and the associated sequence of minimizers of $(LME)_{\beta_k}$, $\{\boldsymbol{p}_{\beta_k}\}_{k\in\mathbb{N}} \subset \mathcal{P}(X)$. Since $\mathcal{P}(X)$ is compact, this sequence has at least a convergent subsequence $\{\boldsymbol{p}_{\beta_{k_j}}\}_{j\in\mathbb{N}}$ converging to $\boldsymbol{q} \in \mathcal{P}(X)$. By uniform convergence and the continuity of f_{β_0} in $\mathcal{P}(X)$, it easily follows that

$$\lim_{j\to\infty} f_{\beta_{k_j}}(\boldsymbol{p}_{\beta_{k_j}}) = f_{\beta_0}(\boldsymbol{q}).$$

From the arguments above, it follows that $f_{\beta_0}(\boldsymbol{p}_{\beta_0}) = f_{\beta_0}(\boldsymbol{q})$, and consequently $\boldsymbol{q} = \boldsymbol{p}_{\beta_0}$, by the uniqueness of the minimizer of f_{β_0}. Thus, any convergent subsequence of $\{\boldsymbol{p}_{\beta_k}\}_{k\in\mathbb{N}}$ converges to $\boldsymbol{p}_{\beta_0}$, and, invoking the compactness of $\mathcal{P}(X)$, we conclude that

$$\lim_{k\to\infty} \boldsymbol{p}_{\beta_k} = \boldsymbol{p}_{\beta_0}.$$

Since the sequence $\{\beta_k\}_{k\in\mathbb{N}}$ converging to β_0 is arbitrary, the continuity of the minimizer as a function of β at any point $\beta_0 \in [0,+\infty)$ immediately follows. □

It follows as an immediate corollary that for $\boldsymbol{x} \in \mathrm{conv}X$,

$$\lim_{\beta\to 0} \boldsymbol{p}_\beta(\boldsymbol{x}) = \boldsymbol{p}_0(\boldsymbol{x}),$$

and thus the *max-ent* convex approximation schemes are recovered from their local counterparts in the limit $\beta \to 0$. Furthermore, since δ in Eq. (23) is independent of $\boldsymbol{x}$, it follows that if $\beta \to \beta_0$ uniformly on $\mathrm{conv}X$ (for instance if the thermalization parameter is uniform in $\mathrm{conv}X$), then $\boldsymbol{p}_\beta \to \boldsymbol{p}_{\beta_0}$ uniformly on $\mathrm{conv}X$.

Finally, we note the following smoothness property $\boldsymbol{p}_\beta$:

Proposition 4.4. *Let $\boldsymbol{x} \in \mathrm{conv}X$. Then, $\boldsymbol{p}_\beta(\boldsymbol{x})$ is a C^∞ function of β in $(0,\infty)$.*

Proof. The proof is identical to that of proposition 4.2 with F regarded as a function of β and $\boldsymbol{\lambda}$. □

4.4.2. Athermal limit The limit $\beta \to +\infty$ is better analyzed by considering the problem equivalent to $(LME)_\beta$ of minimizing the function

$$\widehat{f}_\beta(\boldsymbol{x},\boldsymbol{p}) = U(\boldsymbol{x},\boldsymbol{p}) - \beta^{-1}H(\boldsymbol{p})$$

in $\mathcal{P}_{\boldsymbol{x}}(X)$. The case $\beta = +\infty$ coincides with Rajan's program (RAJ). The methods used in proposition 4.3 only provide a definite answer about the limit of $\boldsymbol{p}_\beta$ as $\beta \to +\infty$ when the points are in general positions:

Proposition 4.5. *Let $\boldsymbol{x} \in \mathrm{conv}X$. Consider a sequence of nonnegative reals $\{\beta_k\}_{k\in\mathbb{N}}$ diverging to $+\infty$ as $k \to +\infty$. Then, every convergent subsequence of $\{\boldsymbol{p}_{\beta_k}(\boldsymbol{x})\}_{k\in\mathbb{N}}$ converges to a solution of (RAJ). Furthermore, if the nodes of X are in general positions, then $\boldsymbol{p}_\beta(\boldsymbol{x})$ converges to the unique solution of (RAJ), the Delaunay convex approximants, as $\beta \to \infty$.*

Proof. The proof is analogous to that of Proposition 4.3. The last argument requires the uniqueness of the limit problem. One simply needs to verify that $\widehat{f}_\beta \to U$ uniformly as $\beta \to \infty$. Indeed, since $\sum_a p_a \log p_a \leq \log N$ for all $\boldsymbol{p} \in \mathbb{R}^N_+$ such that $\mathbf{1}\cdot\boldsymbol{p} = 1$, it immediately follows that $|\widehat{f}_\beta(\boldsymbol{x},\boldsymbol{p}) - U(\boldsymbol{x},\boldsymbol{p})| \leq (\log N)/\beta,\ \forall \boldsymbol{x} \in \mathrm{conv}X,\ \forall \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{x}}(X)$. □

We next show that regardless of the uniqueness of the minimizers of the athermal problem (RAJ), the limit of the minimizers of $(LME)_\beta$ exists as $\beta \to +\infty$. Thus, the *max-ent* regularization of Rajan's program selects a distinguished element of the set of solutions of (RAJ), that which is Pareto optimal, which is unique by virtue of the strict concavity of the entropy in the convex set $\mathcal{S}^{RAJ}_{\boldsymbol{x}}(X)$.

Proposition 4.6. *Let $\boldsymbol{x} \in \mathrm{conv}X$, and consider the solution of (RAJ) with maximum entropy:*

$$\boldsymbol{p}^{Pareto}_\infty(\boldsymbol{x}) = \arg \max_{\boldsymbol{p}\in\mathcal{S}^{RAJ}_{\boldsymbol{x}}(X)} H(\boldsymbol{p}).$$

Then,

$$\lim_{\beta\to+\infty} \boldsymbol{p}_\beta(\boldsymbol{x}) = \boldsymbol{p}^{Pareto}_\infty(\boldsymbol{x}).$$

Proof. We fix $\boldsymbol{x}$ and omit it from the proof. Let $\overline{\boldsymbol{p}}_\infty$ be an arbitrary element of $\mathcal{S}^{RAJ}(X)$ so that $U(\overline{\boldsymbol{p}}_\infty) = m_\infty$, the minimum value of U in $\mathcal{P}(X)$. Let $\{\beta_k\}_{k\in\mathbb{N}}$ be a sequence of nonnegative reals diverging to $+\infty$ as $k \to +\infty$, and $\{\boldsymbol{p}_{\beta_k}\}_{k\in\mathbb{N}}$ the associated sequence of solutions of $(LME)_{\beta_k}$. Let $\{\boldsymbol{p}_{\beta_{k_j}}\}_{j\in\mathbb{N}}$ be a convergent subsequence converging to $\widehat{\boldsymbol{p}}_\infty \in \mathcal{S}^{RAJ}(X)$ according to the preceding proposition. For j large enough, $\beta_{k_j} > 1$, and consequently, by the convexity of $\mathcal{S}^{RAJ}(X)$, we have

$$\boldsymbol{p}_{\infty j} \equiv \left(1 - \frac{1}{\beta_{k_j}}\right)\overline{\boldsymbol{p}}_\infty + \frac{1}{\beta_{k_j}}\widehat{\boldsymbol{p}}_\infty \in \mathcal{S}^{RAJ}(X).$$

Consider also the sequence in $\mathcal{P}(X)$ defined by

$$\boldsymbol{p}_j \equiv \left(1 - \frac{1}{\beta_{k_j}}\right) \boldsymbol{p}_{\infty j} + \frac{1}{\beta_{k_j}} \boldsymbol{p}_{\beta_{k_j}}.$$

It is clear that $\lim_{j\to\infty} \boldsymbol{p}_j = \overline{\boldsymbol{p}}_\infty$. By optimality, we have

$$U(\boldsymbol{p}_{\beta_{k_j}}) - \frac{1}{\beta_{k_j}} H(\boldsymbol{p}_{\beta_{k_j}}) \leq U(\boldsymbol{p}_j) - \frac{1}{\beta_{k_j}} H(\boldsymbol{p}_j),$$

and since

$$U(\boldsymbol{p}_j) = \left(1 - \frac{1}{\beta_{k_j}}\right) m_\infty + \frac{1}{\beta_{k_j}} U(\boldsymbol{p}_{\beta_{k_j}}),$$

it follows that

$$\beta_{k_j} \left(U(\boldsymbol{p}_{\beta_{k_j}}) - m_\infty\right) - H(\boldsymbol{p}_{\beta_{k_j}}) \leq \left(U(\boldsymbol{p}_{\beta_{k_j}}) - m_\infty\right) - H\left(\boldsymbol{p}_j\right). \tag{24}$$

By the continuity of U, $\lim_{j\to\infty} \left(U(\boldsymbol{p}_{\beta_{k_j}}) - m_\infty\right) = 0$. On the other hand, noting that $\widehat{f}_{\beta_{k_j}} \geq U$, we have

$$0 \leq \widehat{f}_{\beta_{k_j}}(\boldsymbol{p}_{\beta_{k_j}}) - U(\widehat{\boldsymbol{p}}_\infty) \leq \widehat{f}_{\beta_{k_j}}(\widehat{\boldsymbol{p}}_\infty) - U(\widehat{\boldsymbol{p}}_\infty) = -\frac{H(\widehat{\boldsymbol{p}}_\infty)}{\beta_{k_j}},$$

and consequently

$$0 \leq \beta_{k_j} \left(U(\boldsymbol{p}_{\beta_{k_j}}) - m_\infty\right) \leq H(\boldsymbol{p}_{\beta_{k_j}}) - H(\widehat{\boldsymbol{p}}_\infty),$$

which by virtue of the continuity of H implies that $\lim_{j\to\infty} \beta_{k_j} \left(U(\boldsymbol{p}_{\beta_{k_j}}) - m_\infty\right) = 0$. Thus, taking limits at both sides of Eq. (24), we conclude that

$$-H(\widehat{\boldsymbol{p}}_\infty) \leq -H(\overline{\boldsymbol{p}}_\infty).$$

Since $\overline{\boldsymbol{p}}_\infty$ is an arbitrary element of $\mathcal{S}^{RAJ}(X)$, we conclude that $\widehat{\boldsymbol{p}}_\infty = \boldsymbol{p}_\infty^{Pareto}$. Thus, every convergent subsequence of $\{\boldsymbol{p}_{\beta_k}\}_{k\in\mathbb{N}}$ has $\boldsymbol{p}_\infty^{Pareto}$ as its limit and, by the compactness of $\mathcal{P}(X)$, we conclude that

$$\lim_{k\to\infty} \boldsymbol{p}_{\beta_k} = \boldsymbol{p}_\infty^{Pareto}.$$

Since the sequence $\{\beta_k\}_{k\in\mathbb{N}}$ diverging to $+\infty$ is arbitrary, the proposition follows. □

Example 4.1. Consider a node set consisting of the four corners of the square $\Omega = [0,1] \times [0,1] \subset \mathbb{R}^2$

$$X = \{\ (0,0), (1,0), (1,1), (0,1)\ \}.$$

For this node set, the Delaunay triangulation is not unique and (RAJ) has multiple solutions. Let us define the bilinear shape functions

$$p_1^{bil}(x,y) = (1-x)(1-y), \quad p_2^{bil}(x,y) = x(1-y), \quad p_3^{bil}(x,y) = xy, \quad p_4^{bil}(x,y) = (1-x)y,$$

and note that $\boldsymbol{p}^{bil}(x,y) \in \mathcal{P}_{(x,y)}(X),\ \forall (x,y) \in \Omega$. It is readily verified that in this particular case (actually for any rectangular configuration) $\boldsymbol{p}^{bil} = \boldsymbol{p}_\beta$ irrespective of the value of β. Indeed, from the first three Kuhn-Tucker conditions

$$0 = \beta[(x-x_a)^2 + (y-y_a)^2] + \log p_a^{bil} + 1 + \lambda_0^* + \lambda_x^*(x_a - x) + \lambda_y^*(y_a - y),\ a = 1,2,3$$

we can obtain explicitly

$$\begin{aligned}
\lambda_0^* &= 1 + \beta x(x-1) + \beta y(y-1) + x\log\frac{1-x}{x} + y\log\frac{1-y}{y} \\
\lambda_x^* &= \beta(2x-1) + \log\frac{1-x}{x} \\
\lambda_y^* &= \beta(2y-1) + \log\frac{1-y}{y}.
\end{aligned}$$

With these Lagrange multiplier values, the fourth Kuhn-Tucker condition is identically satisfied.

Thus, for the square the Pareto set collapses to a single point and, consequently, its $\boldsymbol{p}_\infty^{Pareto} = \boldsymbol{p}^{bil}$, which, according to the propositions above, is the solution of (RAJ) of maximum entropy. This simple examples thus shows that, in addition to the Rajan convex approximants corresponding to the two Delaunay triangulations and their convex combinations, the bilinear shape functions are also a Rajan convex scheme, and in fact the distinguished one. A simple calculation shows that indeed $U = x(1-x) + y(1-y)$ for the bilinear shape functions and for both sets of shape functions associated with the Delaunay triangulations. It is also easy to verify explicitly that the entropy of the bilinear shape functions is greater. This example illustrates how the local *max-ent* approximants eliminate the degeneracy of the Delaunay triangulation.

This example also illustrates the lower semi-continuity of the minimum negative entropy of all the Rajan approximants when seen as a function of the node positions, i. e., of $-H_m(X,\boldsymbol{x}) \equiv -H\left(\boldsymbol{p}_\infty^{Pareto}(\boldsymbol{x})\right)$. Here, $\boldsymbol{p}_\infty^{Pareto}$ corresponds to the node set X. Thus, at degenerate configurations $H_m(X,\boldsymbol{x})$ fails to be continuous with respect to the first argument in general, but is lower-semicontinuous. Consider for instance the following one-parameter family of node sets:

$$X_s = \{\ (-s,-s), (1,0), (1+s,1+s), (0,1)\ \},$$

and consider the average value of the negative entropy of the distinguished Rajan approximant (for $s \neq 0$ there is only one Rajan approximant, the Delaunay approximant) in $\mathrm{conv}X_s$:

$$h(s) = \frac{\int_{\mathrm{conv}X_s} -H_m(X_s,\boldsymbol{x})\ dx}{\mathrm{volume}(\mathrm{conv}X_s)}.$$

It is readily verified that

$$h(s) = \begin{cases} -1 & \text{for } s = 0 \\ -5/6 & \text{otherwise} \end{cases}$$

which demonstrates the lower-semicontinuity of $-H_m(X_s,\boldsymbol{x})$ at $s = 0$. □

Example 4.2. To illustrate the selection of a distinguished Rajan convex approximant by the *max-ent* regularization in degenerate cases, we consider a node set in $\mathbb{R}^2$ containing four co-spherical nodes. The local *max-ent* shape function corresponding to one of the co-spherical

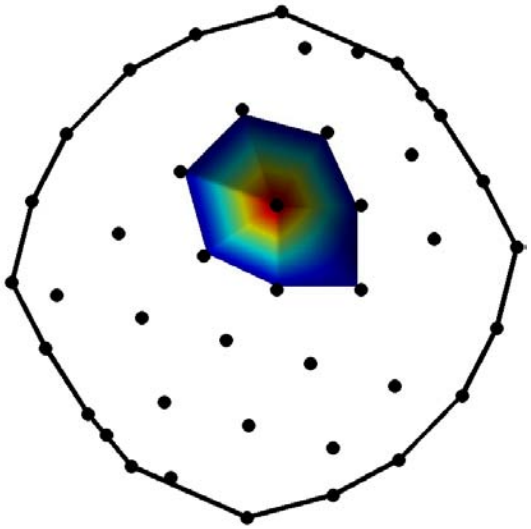


Figure 3. Illustration of the thermal regularization in degenerate cases for $\beta = 10h^{-2}$, where h is the nodal spacing. The nodal set contains four co-spherical nodes that are connected by bilinear shape functions. The remaining nodes are in general positions and are connected by ostensibly affine shape functions

nodes is depicted in Fig. 3 for a value of $\beta = 10h^{-2}$ close to the athermal limit, where h is a representative nodal spacing. It is verified from the figure that the regularization connects the co-spherical nodes with bilinear shape functions, whereas the nodes in general positions are connected by means of ostensibly affine shape functions. □

4.5. Alternative interpretations of the local max-ent *program*

The following interpretations of the local *max-ent* program provide useful insights into the nature of the resulting approximation schemes.

4.5.1. Regularization of the Delaunay program The local *max-ent* program may be regarded as a perturbation of Rajan's program (RAJ). The perturbation regularizes that linear program, possibly having multiple optima, into a one-parameter family of better-behaved smooth and strictly convex programs. The last proposition shows that local *max-ent* approximation effectively removes the degeneracy of the Delaunay triangulation. Thus, when the Delaunay triangulation is not unique the *optimal path* of local *max-ent* approximants converges to a unique distinguished set of Rajan shape functions in the limit. This distinguished shape functions are those of maximum entropy and are Pareto optimal. Thus, the proposed regularization is in analogy to barrier and penalty methods in linear and convex programming, and viscosity solutions of variational problems [25, 26].

4.5.2. A dual regularization of the Delaunay program Alternatively, one can start from the dual (RAJ) problem, an unconstrained non-smooth piecewise-linear convex program (the

minimization of a polyhedral convex function), and approximate it by a one-parameter family of smooth, strictly convex programs. The dual (RAJ) program is

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \max_{a=1,\dots,N} \left(\boldsymbol{\mu}\cdot(\boldsymbol{x}-\boldsymbol{x}_a) - |\boldsymbol{x}-\boldsymbol{x}_a|^2\right), \tag{25}$$

Indeed, equivalent forms of this problem are [24]

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \boldsymbol{\mu}\cdot(\boldsymbol{x}-\boldsymbol{x}_a) - |\boldsymbol{x}-\boldsymbol{x}_a|^2 \le t \quad a=1,\dots,N \end{aligned}$$

and

$$\begin{aligned} \text{maximize} \quad & -\boldsymbol{b}\cdot\boldsymbol{\nu} \\ \text{subject to} \quad & \boldsymbol{A}^T\boldsymbol{\nu} + \boldsymbol{c} \succeq 0 \end{aligned}$$

where we write

$$\boldsymbol{\nu} = \begin{Bmatrix} t \\ \boldsymbol{\mu} \end{Bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} \mathbf{1}^T \\ \boldsymbol{Y} \end{bmatrix}, \quad \boldsymbol{b} = \begin{Bmatrix} 1 \\ \mathbf{0} \end{Bmatrix}, \quad \boldsymbol{c} = \left\{|\boldsymbol{x}-\boldsymbol{x}_1|^2, \dots, |\boldsymbol{x}-\boldsymbol{x}_N|^2\right\}^T. \tag{26}$$

The dual is this program is

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}\cdot\boldsymbol{p} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{p} = \boldsymbol{b} \\ & \boldsymbol{p} \succeq 0 \end{aligned}$$

which is a restatement of (RAJ), as expected. Now consider the *log-sum-exp* function

$$\boldsymbol{lse}(z_1,\dots,z_N) = \log\left(\sum_{a=1}^N \exp z_a\right), \tag{27}$$

and the one-parameter family of analytic functions

$$h_\beta(z_1,\dots,z_N) = \frac{1}{\beta}\boldsymbol{lse}(\beta z_1,\dots,\beta z_N). \tag{28}$$

which, in view of the estimates,

$$0 \le h_\beta(z_1,\dots,z_N) - \max_{a=1,\dots,N} z_a \le \frac{\log N}{\beta}. \tag{29}$$

approximate uniformly the max function as $\beta\to\infty$ [24]. Replacing the max function in Eq. (25) by this approximation gives

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \frac{1}{\beta}\log\left\{\sum_{a=1}^N \exp\left[\beta\boldsymbol{\mu}\cdot(\boldsymbol{x}-\boldsymbol{x}_a) - \beta|\boldsymbol{x}-\boldsymbol{x}_a|^2\right]\right\}, \tag{30}$$

which is a statement of the dual program of $(LME)_\beta$.

4.5.3. Statistical mechanics interpretation The parallel between the local *max-ent* program and statistical mechanics is apparent in the preceding developments and can be further formalized as follows. Consider a system whose configuration space is the index set $I = \{1, \dots, N\}$, and we wish to describe the statistics of this system, i. e. its probability distribution $\boldsymbol{p}$. The energy of configuration $a \in I$ is

$$E_a = |\boldsymbol{x}_a - \boldsymbol{x}|^2 \tag{31}$$

where $\boldsymbol{x}$ is a controllable parameter. Its statistical expectation is the internal energy of the system $U(\boldsymbol{x}, \boldsymbol{p}) = \sum_{a=1}^{N} p_a E_a$. The canonical distribution of the system then follows as the solution to the variational problem

$$\inf_{\boldsymbol{p} \in \mathcal{P}_{\boldsymbol{x}}(X)} \left\{ U(\boldsymbol{x}, \boldsymbol{p}) - \beta^{-1} H(\boldsymbol{p}) \right\} \tag{32}$$

where $\beta = 1/k_B T$, k_B is Bolzmann's constant and T is the absolute temperature. Evidently, (32) is just a re-statement of the problem $(LME)_\beta$. The value of the infimum $F(\boldsymbol{x}, \beta)$ in (32) is the free energy of the system. From this perspective, the local *max-ent* problem $(LME)_\beta$ may be regarded as a *thermalization* of the Rajan problem (RAJ). Conversely, the Rajan problem, may be regarded as the zero temperature limit of the thermalized problem. Thus, thermalization replaces a problem of energy minimization at zero temperature, namely the Rajan problem (RAJ), by the problem of computing the partition function (16) at finite temperature. Since the configuration space is the index set I, the evaluation of the partition function reduces to the computation of a finite sum. In this manner, the solution of the linear programming problem (RAJ) is replaced by an explicit calculation. This connection between the computation of partition functions and energy minimization in the limit of zero temperature indeed was the original insight that led to the present work.

4.6. Examples

Figure 4 shows the local *max-ent* shape function and its partial derivatives for a node in a two-dimensional node set as a function of the dimensionless parameter $\gamma = \beta h^2$, where h is a measure of the nodal spacing and β is constant over the domain. It can be seen from this figure that the shape functions are smooth and their degree of locality is controlled by the parameter γ. For the maximum value of $\gamma = 6.8$ shown in the figure the shape function ostensibly coincides with the Delaunay shape function.

The parameter β can be allowed to depend on position and that dependence can be adjusted adaptively in order to achieve varying degrees of locality. Figure 5 shows an illustration of this type of adaptivity. In this example, the function $\beta(x)$ is chosen such that the finite element limit is attained at the left and center nodes of the node set, with increasing thermalization away from these nodes. In applications, the question of the optimal distribution of β over the domain of analysis arises naturally. In problems whose solutions minimize a certain functional, a natural strategy is to select $\beta(x)$ *variationally*, i. e., to let $\beta(x)$ be such function as minimizes the functional. However, this enhancement of the method will not be pursued in this paper.

Figure 6(a) illustrates the behavior of the local *max-ent* shape functions at the boundary of the domain. In particular it should be noted that the shape functions of interior nodes vanish at the boundary, and that the shape functions of extreme nodes equal one at their corresponding node.

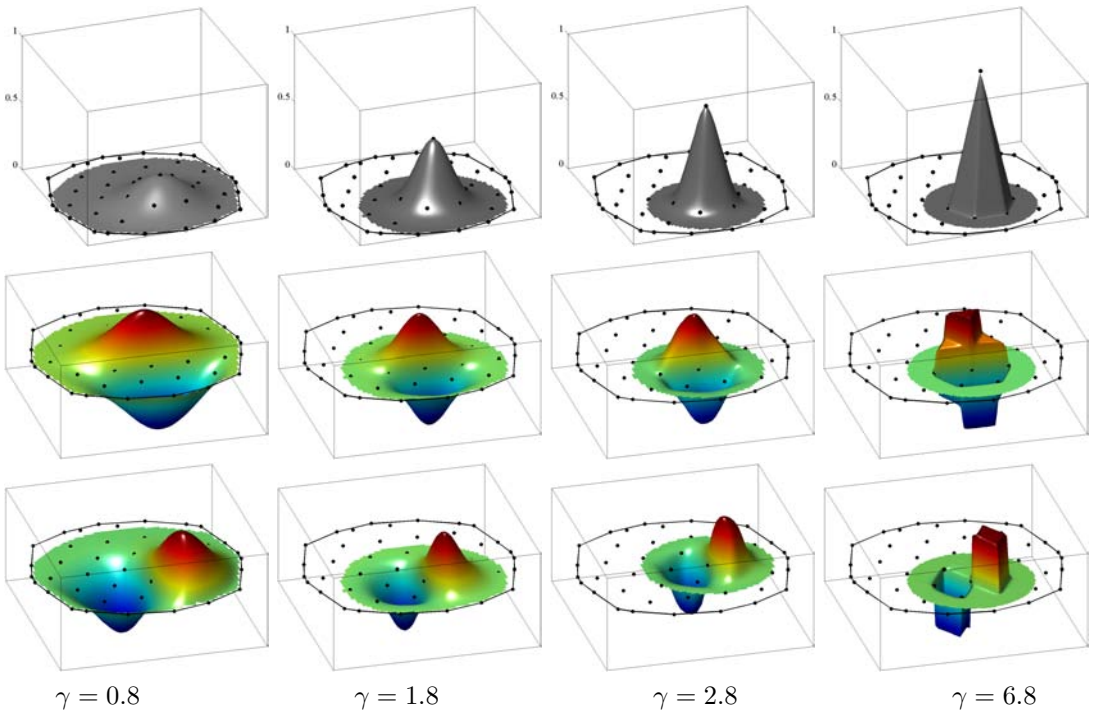


Figure 4. Local *max-ent* shape functions for a two dimensional arrangement of nodes, and spacial derivatives (arbitrary scale) for several values of $\gamma = \beta h^2$.

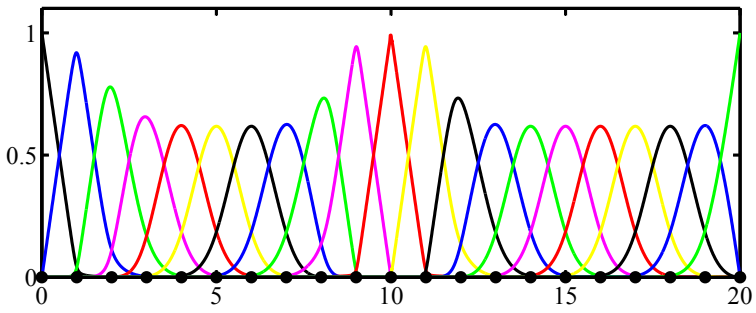
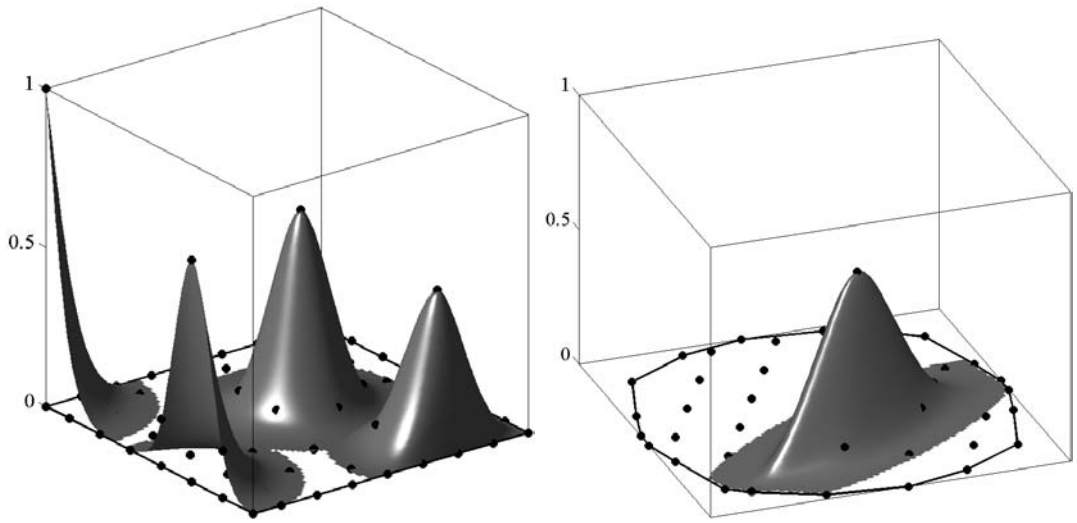


Figure 5. One dimensional example of seamless transition to finite elements achieved by tuning the function $\beta(x)$.



(a) (b)

Figure 6. (a) Illustration of the behavior of the local *max-ent* shape functions at the boundary of the domain. (b) Example of anisotropic local *max-ent* shape function.

Finally, we briefly discuss measures of locality other than (11). For instance, suppose that we adopt a general distance d in place of the standard Euclidean distance employed in (11). The width of shape function p_a is then

$$w[p_a] = \int_\Omega p_a(\boldsymbol{x}) d^2(\boldsymbol{x}, \boldsymbol{x}_a)\ dx. \tag{33}$$

In this case, the local *max-ent* shape functions become

$$p_a(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x}, \boldsymbol{\lambda}^*(\boldsymbol{x}))} \exp\left[-\beta(\boldsymbol{x}) d^2(\boldsymbol{x}, \boldsymbol{x}_a) + \boldsymbol{\lambda}^*(\boldsymbol{x}) \cdot (\boldsymbol{x} - \boldsymbol{x}_a)\right]$$

with the partition function appropriately modified. In particular, this extension can be useful for purposes of defining anisotropic shape functions, e. g., in problems involving localization. For instance, Fig. 6(b) shows a shape function defined using an Euclidian distance of the form $d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y}) \cdot \boldsymbol{G}(\boldsymbol{x} - \boldsymbol{y})}$, for a constant metric tensor $\boldsymbol{G}$. The localization of the shape function to a plane is noteworthy in the figure.

Another variation of the standard local *max-ent* program is to consider measures of locality that are not quadratic in the distance. In this manner, shape functions with different decay behavior may be obtained. For instance, consider width of the form

$$w[p_a] = \int_\Omega p_a(\boldsymbol{x}) |\boldsymbol{x} - \boldsymbol{x}_a|^p\ dx. \tag{34}$$

where $1 \leq p \leq \infty$. Fig. 7(a)-(c) shows the resulting shape functions for $p = 1, 3, 6$ and $\gamma = 2.8, 1.4, 1.0$, respectively. It is interesting to note that the l^1-shape functions are not differentiable at their corresponding nodes. Fig. 7(d) shows the function resulting from the

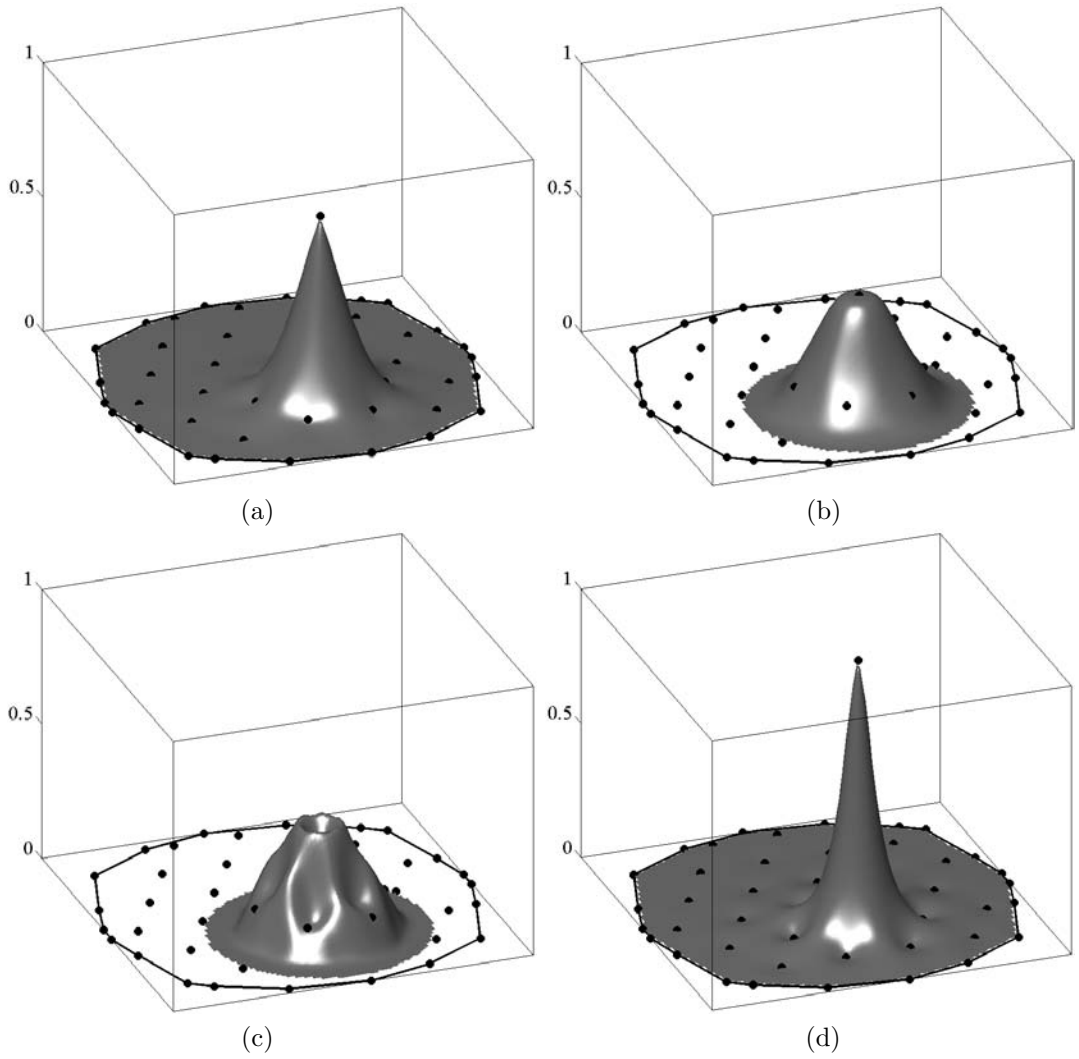


Figure 7. Local *max-ent* shape functions for l^p-distances. a) $p = 1$ and $\gamma = 2.8$; b) $p = 3$ and $\gamma = 1.4$; c) $p = 6$ and $\gamma = 1.0$; d) log function and $\gamma = 2.0$.

function $\log|\boldsymbol{x} - \boldsymbol{x}_a|$ with $\gamma = 2$, which leads to a decay of the form $\sim r^{-2}$ with distance r. Interestingly, shape functions constructed from this locality measure verify the Kronecker-delta property.

4.7. Practical evaluation of the shape functions

In practice, the evaluation of the local *max-ent* approximants at a given point $\boldsymbol{x} \in \text{conv}X$ does not require the solution of the $(LME)_\beta$ as a constrained convex program involving N

Table I. Average number of Newton-Raphson iterations per sample point.

γ	Tol_{NR}		
	10^{-5}	10^{-10}	machine precision
0.8	2.7	3.6	4.2
1.8	2.4	3.5	4.3
2.8	3.0	4.0	5.0
6.8	4.0	5.0	6.0

unknowns. Instead, as shown in § 4.2, it is sufficient to solve an unconstrained minimization problem, Eq. (18), involving at most d unknowns, namely, the face dimension of $\boldsymbol{x}$. This problem is smooth, strictly convex and, by virtue of the Kuhn-Tucker theorem, is guaranteed to have a unique solution. This fact confers efficiency and robustness to the calculation of the local *max-ent* approximants. The resulting set of nonlinear equations

$$\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda}) = 0 \tag{35}$$

for $\boldsymbol{\lambda}^*$, where $\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda})$ is given by Eq. (19), can be solved numerically, e. g., by means of a few Newton-Raphson interation. The requisite Hessian matrix is given by Eq. (20).

To illustrate the effectiveness of the Newton-Raphson method to solve Eq. (35), the average number of iterations required for convergence with the criterion $|\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda}^k)| \leq \mathrm{Tol}_{NR}$ is presented in Table I for different values of γ and different tolerances. The node set of Fig. 4 is used in this example, and the fine grid of sample points used to generate these plots is used to compute the average. It is observed that, for a tolerance of 10^{-5}, between 2 and 3 iterations are sufficient over a broad range of values of the thermalization parameter. It is also observed that, as the athermal limit is reached (for large values of γ), the Newton-Raphson method needs more iterations for convergence. This slowdown in convergence is expected since the objective function tends to a faceted polyhedral convex function as $\beta \to +\infty$ (cf §4.5.2).

It is evident from Eq. (17) that the support of the shape functions is the entire domain $\mathrm{conv}X$, i. e., the local *max-ent* shape functions have global support. However, the shape functions decay as $\exp(-\beta\, r^2)$ with distance r to the corresponding node, and can thus be truncated at distances greater than a small multiple of $\beta^{-1/2}$. This decay property establishes a connection with Gaussian radial basis functions (cf, e. g., [4]) and has the important consequence that only a small number of nodes contribute ostensibly to the partition function, which greatly reduces the computational cost of the solution of problem (18). For practical purposes, a tolerance Tol_0 is set below which we set the shape functions to zero. Therefore, owing to the Gaussian decay, the numerically effective support of the shape function corresponding to node $\boldsymbol{x}_a$ is the ball centered at this point of radius $R_a = \sqrt{-\log(\mathrm{Tol}_0)/\beta} = h\,\sqrt{-\log(\mathrm{Tol}_0)/\gamma}$. By way of illustration, if the tolerance is the double precision machine precision ($\sim 10^{-16}$) and $\gamma = 1.8$, then the radius is below 4.5 times the nodal spacing. The fast decay of the shape functions can be observed in the first three columns of Fig. 4, where the support has been determined for a tolerance of $\mathrm{Tol}_0 = 10^{-6}$.

Once the shape functions have been evaluated, the calculation of the derivatives of the shape functions is explicit by means of Eq. (21).

5. APPLICATION TO LINEAR AND NONLINEAR ELASTICITY

Local *max-ent* approximation schemes provide a convenient choice of basis to use in conjunction with the Rayleigh-Ritz—or constrained energy minimization—approach to elasticity problems. In this section, we first present a displacement patch-test, and then proceed to demonstrate the accuracy and convergence characteristics of the local *max-ent* solutions by means of two examples of application: the standard benchmark problem of a linear elastic built-in cantilever beam loaded at the tip; and the upsetting and extension of a block of compressible neo-Hookean rubber. In all calculations we confine our attention to uniformly distributed node sets and the parameter β is taken to be uniform. In addition, all numerical integrals are carried out using standard quadrature rules based on the limiting Delaunay triangulation. For large values of β, the local *max-ent* approximation scheme differs little from simplicial interpolation and a one-point quadrature rule is found to suffice. For lower values of β a three point rule for triangles and a four point rule for tetrahedra are used. The accuracy of the local *max-ent* solutions is compared to that of the Delaunay linear finite elements. This comparison is natural in the present context since linear finite elements arise in the athermal limit.

5.1. Patch test

In the displacement patch test, the boundary of the computational domain is subjected to an affine transformation. For the numerical method to pass the test, the numerical solution in the interior of the domain must reproduce this affine transformation exactly. Since the local *max-ent* approximants satisfy the first order reproducing condition, the patch test is passed to machine precision if exact integration is used. Thus, the test assesses the numerical quadrature and the accuracy in the calculation of the shape functions.

Consider the square $[0,1]\times[0,1]$ of a linear isotropic elastic material characterized by Young modulus $E = 1$ and Poisson's ratio $\nu = 0.3$. The boundary of the square is subjected to a linear transformation characterized by the matrix

$$\begin{pmatrix} 1 & -\sqrt{3}/2 \\ \sqrt{3} & 1/2 \end{pmatrix}.$$

The two node sets depicted in Fig. 8 are considered. The figure also shows a typical quadrature rule, based on the underlying Delaunay triangulation. Symmetric rules ranging form 1 to 175 points per triangle are considered in the calculations [27].

Table II reports the relative L_2 errors for two different node sets, for two values of γ, and for nine different quadrature rules. To specifically investigate the influence of numerical quadrature, we have assigned Tol_{NR} and Tol_0 to machine precision in this table. It can be observed that for the structured node set, it is possible to pass the patch test ostensibly within machine precision with 175 integration points per triangle. However, for an unstructured node set the errors reduce more slowly as the number of quadrature points is increased, and a maximum precision of 3×10^{-6} is achieved. These results conform to previous experience for other meshfree methods, special quadrature rules notwithstanding [28]. It should also be mentioned that in the range of one to six quadrature points per triangle there is no significant difference between the structured and unstructured node sets. It is also noteworthy that, for the node set (b), a lower value of γ requires a more accurate quadrature for a given error in the patch test. The relative errors reported in Table II for the first four rules are not affected if

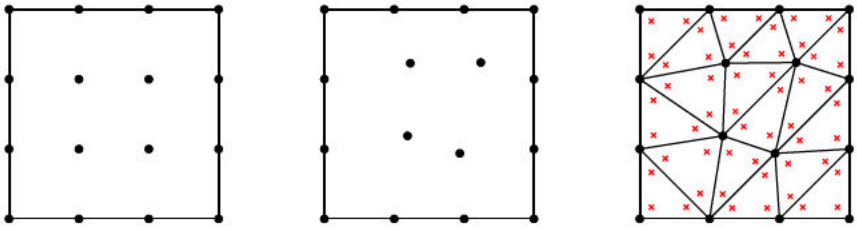


Figure 8. Node sets (a) and (b) used for the patch test, and typical layout of the quadrature points based on the underlying Delaunay triangulation.

Table II. Relative errors in the L_2 norm for the patch test

quadrature	(a) $\gamma = 1.8$	(a) $\gamma = 0.8$	(b) $\gamma = 1.8$	(b) $\gamma = 0.8$
1	3×10^{-3}	1×10^{-2}	2×10^{-3}	1×10^{-2}
3	5×10^{-4}	3×10^{-4}	4×10^{-4}	1×10^{-3}
6	4×10^{-4}	3×10^{-4}	3×10^{-4}	1×10^{-3}
7	3×10^{-5}	6×10^{-6}	2×10^{-4}	2×10^{-4}
12	8×10^{-6}	7×10^{-8}	8×10^{-5}	2×10^{-4}
25	4×10^{-8}	2×10^{-11}	3×10^{-5}	8×10^{-5}
54	3×10^{-10}	7×10^{-16}	1×10^{-5}	3×10^{-5}
126	7×10^{-14}	4×10^{-16}	5×10^{-6}	1×10^{-5}
175	3×10^{-16}	6×10^{-16}	3×10^{-6}	7×10^{-6}

$\mathrm{Tol}_{NR} = \mathrm{Tol}_0 = 10^{-5}$. These are the tolerances used in subsequent examples for computational efficiency.

5.2. Linear elasticity: cantilever beam

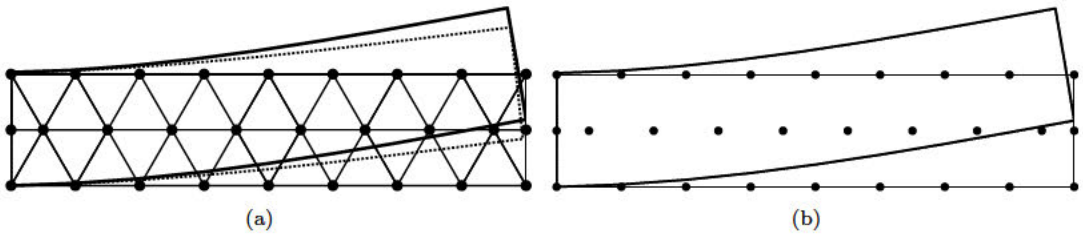


Figure 9. Linear elastic built-in cantilever loaded at the tip, deformed configurations. a) Finite element solution; b) Local *max-ent* solution. Also shown for reference as a solid line is the exact deformed configuration.

We begin by considering the standard benchmark problem of a linear elastic cantilever beam acted upon by a parabolic distribution of tractions at one end and built-in boundary conditions at the other end. The exact solution to this problem is known and may be looked up in [29]. We consider beams whose length-to-height ratio is $8/\sqrt{3}$. The problem is symmetric with respect

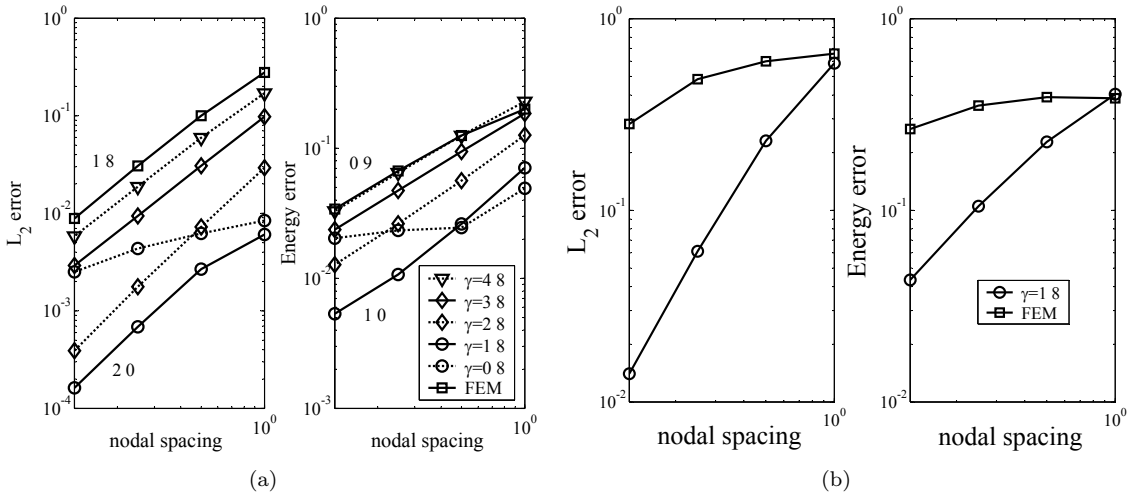


Figure 10. Linear elastic built-in cantilever loaded at the tip, convergence plots. (a) $\nu = 0.3$ and (b) $\nu = 0.499$.

to the neutral axis of the beam, and therefore only the upper half of the beam is analyzed.

It should be noted that, owing to the weak Kronecker-delta property of the local *max-ent* approximants, the imposition of Dirichlet boundary conditions is straightforward without degrading the optimal rate of convergence: it is sufficient to set the nodal displacement to the prescribed values. Indeed, owing to the properties outlined in Section 2.2, the basis functions corresponding to the nodes on a face alone retain the full approximation properties on that face.

Figure 9 compares the exact deformed configuration of the beam and the numerical deformed configurations obtained by means of linear finite elements and local *max-ent* shape functions with $\gamma = 1.8$. In this example, Poisson's ratio is $\nu = 0.3$ and thus volumetric locking is not an issue. The superior accuracy of the local *max-ent* approximation scheme is clearly apparent from this figure.

Convergence plots are shown Fig. 10. In this plots, the L_2 norm of the error is defined as

$$\|\boldsymbol{e}\|_{L_2} = \left(\int_\Omega |\boldsymbol{u} - \boldsymbol{u}^h|^2 \, d\boldsymbol{x} \right)^{1/2}, \tag{36}$$

whereas the energy semi-norm is

$$\|\boldsymbol{e}\|_E = \left(\frac{1}{2} \int_\Omega (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^h) : (\boldsymbol{\sigma} - \boldsymbol{\sigma}^h) \, d\boldsymbol{x} \right)^{1/2}, \tag{37}$$

In these expressions $\boldsymbol{u}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\sigma}$ denote the exact displacement, strain and stress fields, and $\boldsymbol{u}^h$, $\boldsymbol{\varepsilon}^h$, and $\boldsymbol{\sigma}^h$ denote the corresponding numerical approximations. The convergence plots display the numerical errors normalized by the norm of the exact solution as a function of a normalized nodal spacing.

Fig. 10(a) shows the convergence plots for the case of $\nu = 0.3$. It is evident from these plots that there is an optimal value for the parameter γ of around 1.8 for which accuracy is maximized. A similar behavior is observed in meshfree methods in terms of the dilation parameter. For very low values of γ convergence is degraded. This is due to the numerical quadrature: with the 12 point rule and $\gamma = 0.8$, the optimal rate of convergence is recovered. It should be noted that for the other values of γ considered, improving the numerical quadrature does not affect noticeably the numerical errors, which are mostly due to the approximation properties of the local *max-ent* schemes. As expected, the behavior of the local *max-ent* solutions approaches that of the Delaunay solutions as γ increases. The local *max-ent* solutions exhibit optimal convergence rates, whereas the Delaunay solutions display slightly suboptimal rates. Furthermore, the accuracy of the local *max-ent* solutions is vastly superior to that of the Delaunay solutions, as evidenced by the large shift in the convergence curves. The distinct advantage of the local *max-ent* approximations over Delaunay approximations persists when the cost of the solutions is carefully accounted for.

Figure 10(b) shows the convergence plots for a nearly incompressible material. It is well known that increasing the support size in MLS meshfree methods can alleviate the problem of volumetric locking. However, locking is not completely eliminated (cf [4] and references therein) since the discretization schemes do not commute exactly with the Helmoltz-Hodge decomposition of the displacement field. This type of behavior is also exhibited by local *max-ent* approximations; for nearly incompressible materials, the asymptotic rate of convergence is reached at much coarser discretizations for the local *max-ent* approximants than for simplicial finite elements (cf Fig. 10(b)). Nevertheless, as the incompressible limit is reached locking occurs for both methods.

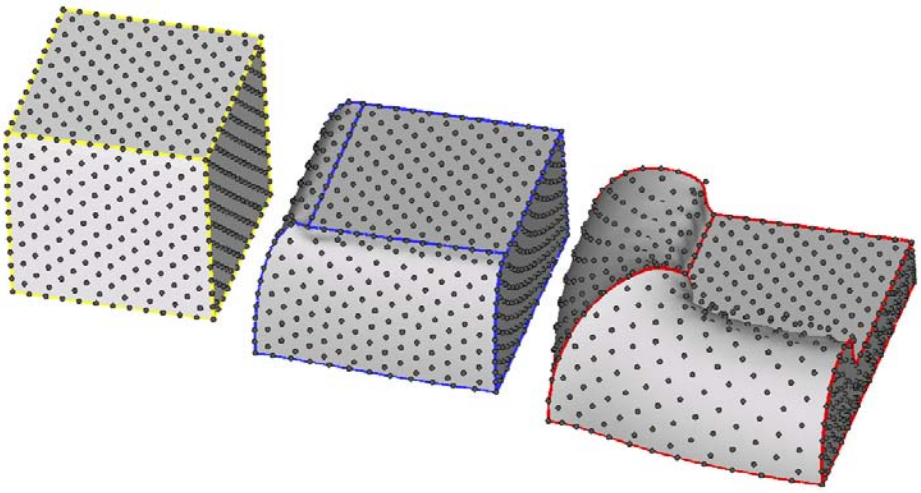


Figure 11. Compression of a hyper-elastic block.

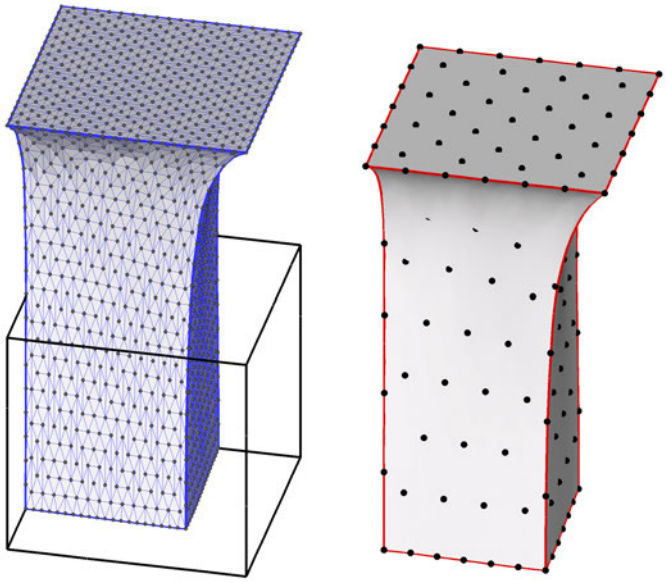


Figure 12. Final deformation for the finest FE mesh and second-coarsest local *max-ent* discretization ($\nu_0 = 0.495$).

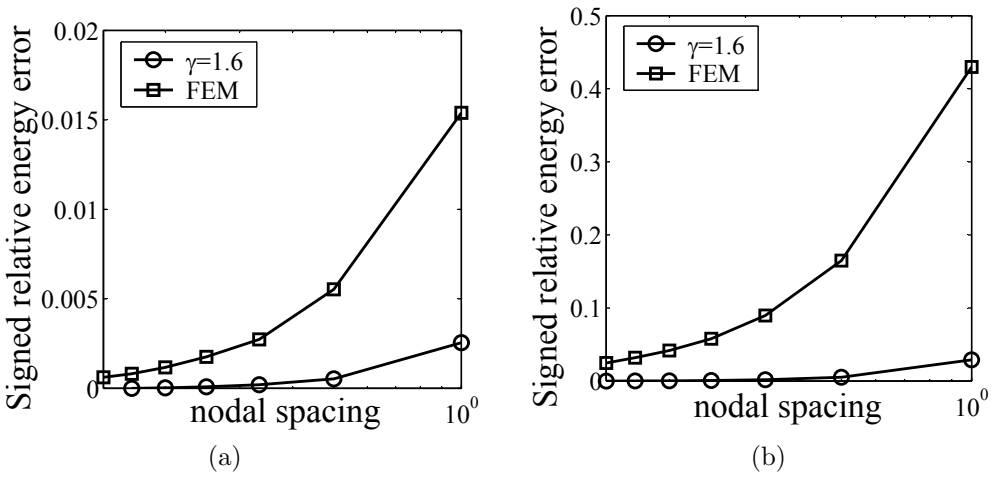


Figure 13. Signed relative error in strain energy with respect to a reference numerical solution for (a) $\nu_0 = 0.333$, and (b) $\nu_0 = 0.495$.

5.3. Nonlinear elasticity: hyper-elastic block

Next we demonstrate the performance of local *max-ent* approximations in strongly nonlinear problems. The test case under consideration concerns the compression and tension of a hyperelastic block. The material is compressible-neo Hookean with energy density

$$W(\boldsymbol{F}) = \frac{1}{2}\lambda \log^2(J) - \mu \log(J) + \frac{\mu}{2}\mathrm{tr}(\boldsymbol{F}^T\boldsymbol{F}) \tag{38}$$

where $\boldsymbol{F}$ is the deformation gradient, $J = \det(\boldsymbol{F})$ and λ and μ are Lamè constants. Two sets of material constants are considered: $\lambda/\mu = 2$, corresponding to an initial Poisson's ratio of $\nu_0 = 0.333$; and $\lambda/\mu = 100$, corresponding to an initial Poisson's ratio of $\nu_0 = 0.495$. By the symmetry of the problem only an eighth of the sample needs to be analyzed. Quasi-static loading conditions are considered. The prescribed displacements are applied incrementally and the total potential energy is minimized by the conjugate gradient method.

Figure 11 shows the deformation of the block at 75% compression for the second parameter set. The robustness of the method for nonlinear problems involving severe deformations is evident from this figure. It is interesting to note that some nodes lie outside the deformed body, which is a consequence of the non-interpolatory character of the shape functions.

Figure 13 shows the deformed configurations of the block at 100% tensile deformation. The calculation is performed with seven uniform node sets of variable resolution and for both sets of material constants. Figure 13 shows the dependence of a normalized signed relative error in strain energy (relative to an overkill numerical solution) on the nodal spacing. It is evident from this figure that the numerically computed potential energy decreases monotonically with mesh refinement, as expected. Figure 13(a) depicts convergence curves for the local *max-ent* and finite element solutions in the compressible case ($\nu_0 = 0.333$). It is observed from that figure that the accuracy of the local *max-ent* solution is vastly superior to that of the finite element solution. The finest finite element solution has a comparable—albeit slightly larger—error than the second-coarsest local *max-ent* solution, Fig. 12. By contrast, the CPU time incurred by the local *max-ent* solution is over a hundred times shorter than that of the finite element solution. This difference in performance is more pronounced in the nearly incompressible case, Fig. 13(b). In this case, the finite element solution converges very slowly and the coarse meshes result in very large errors. Indeed, the coarsest local *max-ent* solution suffices to achieve an accuracy comparable to that of the finest finite element mesh, at a CPU-time knockdown factor of one thousand.

6. SUMMARY AND CONCLUDING REMARKS

We have developed a new type of approximation scheme, which we term local *max-ent* approximation scheme, which represents a compromise between unbiased statistical inference, in the sense of information theory, and the desire to define shape functions of the least width possible. The resulting shape functions are non-negative, possess a weak Kronecker-delta property at the boundary, and reduce in the limit to piecewise affine interpolation over a Delaunay triangulation, whereas away from this limit the shape functions are smooth and analogous to those employed in MLS schemes. In this sense, local *max-ent* approximation supplies an efficient bridge between simplicial finite elements and meshfree methods. In particular, by adjusting the spatial variation $\beta(\boldsymbol{x})$, it is possible to select regions of the domain

of analysis which are treated by finite elements and regions that are treated in the style of meshfree methods, with seamless smooth transitions between those regions. This is in contrast to other approaches coupling meshfree schemes with finite elements (cf [4] and references therein). Although we have not developed here the blending of the local *max-ent* approximants with other convex approximants such as subdivision approximants, NURBS, or linear finite elements, the variational nature of our method makes this coupling straightforward. The calculation of the shape functions can be carried out simply, robustly and efficiently in any spatial dimension. The numerical tests presented in this paper suggest that, for problems possessing smooth solutions, local *max-ent* approximations supply high accuracy at low cost.

We conclude by pointing out some of the limitations of the method and opportunities for further development. As defined in this paper, local *max-ent* shape functions are necessarily positive, which precludes the formulation of high-order schemes satisfying the second order consistency condition in Eq. (9). The positivity constraint is introduced mainly in order enable the interpretation of the shape functions as probability densities, which in turn facilitates the conceptual connections with information theory. Higher-order schemes may be obtained by relaxing the requirement expressed in Eq. (9). These developments, as well as the formulation of strictly compactly supported *max-ent* approximants, will be reported in a subsequent paper.

The local *max-ent* approximation schemes have been defined in the convex hull of the node set. If non-convex domains are treated, the local *max-ent* approximants lose the weak Kronecker-delta property in the non-convex parts of the boundary, and thus behave similarly to MLS approximants. The effective treatment of non-convex domains is of considerable interest and has been extensively studied in the context of MLS-based meshfree methods [30, 31, 32]. Visibility, diffraction, and constrained path criteria have been proposed that modify how nearby nodes interact in the vicinity of a non-convex part of the boundary. These methods are directly applicable to local *max-ent* approximation. For instance, it is possible to replace the Euclidean distance $|\boldsymbol{x} - \boldsymbol{x}_a|$ in the definition of the shape functions by the length of the shortest path contained within the domain connecting $\boldsymbol{x}$ and $\boldsymbol{x}_a$. Alternatively, the non-convex domain can be decomposed into convex sub-domains and approximation schemes can then be constructed separately in each of the sub-domains. The schemes in each sub-domain are guaranteed to be conforming by the *conforming patches* property of local *max-ent* approximation schemes.

Finally we remark briefly on the possibility of adapting the function $\beta(\boldsymbol{x})$. In problems having a variational structure and where the solutions obey a minimum principle, the natural approach is to let the minimum principle itself select the optimal function $\beta(\boldsymbol{x})$. In this view, the energy function is minimized with respect to the displacement field and with respect to $\beta(\boldsymbol{x})$. This program is facilitated by the ability to compute explicitly derivatives of shape functions with respect to β, and by the guaranteed solvability of Eq. (18) for any non-negative value of β. This is in contrast to the problem of choosing an optimal value of the dilation parameter in MLS approximations. In this case, it is often not straightforward to obtain analytical derivatives of the shape functions with respect to the dilation parameter. In addition, the dilation parameter is subject to lower-bound solvability constraints which are difficult to verify *a priori*.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Department of Energy through Caltech's ASCI

ASAP Center for the Simulation of the Dynamic Response of Materials, and the support received from NSF through an ITR grant on Multiscale Modeling and Simulation and Caltech's Center for Integrative Multiscale Modeling and Simulation.

APPENDIX

II. Spacial derivatives of the shape functions

In this appendix we detail the procedure for the calculation of the spatial derivative of the shape functions. For the sake of generality we consider the case in which the parameter $\beta(\boldsymbol{x})$ is spatially non-uniform. We denote spatial gradients of real functions by ∇, whereas for vector-valued functions we denote by $D\boldsymbol{y}(\boldsymbol{x})$ the matrix partial derivatives. The symbol ∂ denotes partial differentiation. We define the following functions

$$f_a(\boldsymbol{x},\boldsymbol{\lambda},\beta) = -\beta|\boldsymbol{x}-\boldsymbol{x}_a|^2 + \boldsymbol{\lambda}\cdot(\boldsymbol{x}-\boldsymbol{x}_a), \tag{39}$$

$$p_a(\boldsymbol{x},\boldsymbol{\lambda},\beta) = \frac{\exp\left[f_a(\boldsymbol{x},\boldsymbol{\lambda},\beta)\right]}{\sum_b \exp\left[f_b(\boldsymbol{x},\boldsymbol{\lambda},\beta)\right]}, \tag{40}$$

$$\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda},\beta) = \sum_a p_a(\boldsymbol{x},\boldsymbol{\lambda},\beta)(\boldsymbol{x}-\boldsymbol{x}_a), \tag{41}$$

$$\boldsymbol{J}(\boldsymbol{x},\boldsymbol{\lambda},\beta) = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\lambda}} = \sum_a p_a(\boldsymbol{x},\boldsymbol{\lambda},\beta)(\boldsymbol{x}-\boldsymbol{x}_a)\otimes(\boldsymbol{x}-\boldsymbol{x}_a) - \boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda},\beta)\otimes\boldsymbol{r}(\boldsymbol{x},\boldsymbol{\lambda},\beta). \tag{42}$$

Given a function $h(\boldsymbol{x},\boldsymbol{\lambda},\beta)$, we define the function h^* that depends only on $\boldsymbol{x}$ as

$$h^*(\boldsymbol{x}) = h\left(\boldsymbol{x},\boldsymbol{\lambda}^*(\boldsymbol{x}),\beta(\boldsymbol{x})\right),$$

where $\boldsymbol{\lambda}^*(\boldsymbol{x})$ is the unique maximizer of

$$g(\boldsymbol{\lambda}) = -\log\left\{\sum_a \exp\left[f_a(\boldsymbol{x},\boldsymbol{\lambda},\beta(\boldsymbol{x})\right]\right\}.$$

Our goal is to compute ∇p_a^*. It is readily verified that

$$\nabla p_a^* = p_a^*\left(\nabla f_a^* - \sum_b p_b^*\nabla f_b^*\right).$$

By the chain rule, we have

$$\nabla f_a^* = \left(\frac{\partial f_a}{\partial \boldsymbol{x}}\right)^* + \left(\frac{\partial f_a}{\partial \boldsymbol{\lambda}}\right)^* D\boldsymbol{\lambda}^* + \left(\frac{\partial f_a}{\partial \beta}\right)^*\nabla\beta, \tag{43}$$

where

$$\left(\frac{\partial f_a}{\partial \boldsymbol{x}}\right)^* = -2\beta\,(\boldsymbol{x}-\boldsymbol{x}_a) + \boldsymbol{\lambda}^*(\boldsymbol{x}), \quad \left(\frac{\partial f_a}{\partial \boldsymbol{\lambda}}\right)^* = (\boldsymbol{x}-\boldsymbol{x}_a), \quad \left(\frac{\partial f_a}{\partial \beta}\right)^* = -|\boldsymbol{x}-\boldsymbol{x}_a|^2.$$

The only term that is not available explicitly in eq. (43) is $D\boldsymbol{\lambda}^*$. In order to compute this term we note that, since $\boldsymbol{r}^*$ is identically zero,

$$0 = D\boldsymbol{r}^* = \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{x}}\right)^* + \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\lambda}}\right)^* D\boldsymbol{\lambda}^* + \left(\frac{\partial \boldsymbol{r}}{\partial \beta}\right)^*\otimes\nabla\beta.$$

Simple calculations show that

$$\left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\lambda}}\right)^* = \boldsymbol{J}^*, \quad \left(\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{x}}\right)^* = -2\beta\ \boldsymbol{J}^* + \boldsymbol{id}, \quad \left(\frac{\partial \boldsymbol{r}}{\partial \beta}\right)^* = -\sum_a p_a^* |\boldsymbol{x} - \boldsymbol{x}_a|^2 (\boldsymbol{x} - \boldsymbol{x}_a),$$

whence it follows that

$$D\boldsymbol{\lambda}^* = 2\beta\ \boldsymbol{id} - (\boldsymbol{J}^*)^{-1} + (\boldsymbol{J}^*)^{-1} \left(\sum_a p_a^* |\boldsymbol{x} - \boldsymbol{x}_a|^2 (\boldsymbol{x} - \boldsymbol{x}_a) \otimes \nabla\beta \right).$$

Rearranging terms, and noting that p_a^* verifies the linear consistency condition, we finally obtain

$$\nabla p_a^* = -p_a^* \left(\boldsymbol{J}^*\right)^{-1} (\boldsymbol{x} - \boldsymbol{x}_a) + p_a^* K_a \nabla\beta, \tag{44}$$

where we write

$$K_a = \left[\sum_b p_b^* |\boldsymbol{x} - \boldsymbol{x}_b|^2 (\boldsymbol{x} - \boldsymbol{x}_b) \right] \cdot \left(\boldsymbol{J}^*\right)^{-1} (\boldsymbol{x} - \boldsymbol{x}_a) - |\boldsymbol{x} - \boldsymbol{x}_a|^2 + U(\boldsymbol{x}, \boldsymbol{p}^*).$$

REFERENCES

1. Nayroles B, Touzot G, Villon P. Generalizing the finite element method: diffuse approximation and diffuse elements. *Computational Mechanics* 1992; **10**(5):307–318.
2. Belytschko T, Lu YY, Gu L. Element-free Galerkin methods. *International Journal for Numerical Methods in Engineering* 1994; **37**(2):229–256.
3. Liu WK, Li S, Belytschko T. Moving least square reproducing kernel methods Part I: Methodology and convergence. *Computer Methods in Applied Mechanics and Engineering* 1997; **143**(1-2):113–154.
4. Huerta A, Belytscko T, Fernández-Méndez S, Rabczuk T. *Encyclopedia of Computational Mechanics*, volume 1, chapter Meshfree methods, pages 279–309. Wiley: Chichester, 2004; .
5. Duarte CA, Oden JT. An h-p adaptive method using clouds. *Computer Methods in Applied Mechanics and Engineering* 1996; **139**(1-4):237–262.
6. Babuška I, Melenk JM. The partition of unity method. *International Journal for Numerical Methods in Engineering* 1997; **40**(4):727–758.
7. Jaynes ET. Information theory and statistical mechanics. *Physical Review* 1957; **106**(4):620–630.
8. Rockafellar RT. *Convex analysis*. Princeton University Press: Princeton, NJ, 1970.
9. Prautzch H, Boehm W, Paluszny M. *Bézier and B-spline techniques*. Springer-Verlag: Berlin, 2002.
10. DeVore RA. *The approximation of continuous functions by positive linear operators*. Springer-Verlag: Berlin, 1972.
11. Sukumar N, Moran B, Belytschko T. The natural element method in solid mechanics. *International Journal for Numerical Methods in Engineering* 1998; **43**(5):839–887.
12. Cirak F, Ortiz M, Schröder P. Subdivision surfaces: a new paradigm for thin-shell finite-element analysis. *International Journal for Numerical Methods in Engineering* 2000; **47**(12):2039–2072.
13. Cottin C, Gavrea I, Gonska HH, Kacsó DP, Zhou DX. Global smoothness preservation and variation-diminishing property. *Journal of Inequalities and Applications* 1999; **4**(2):91–114.
14. Fernández-Méndez S, Huerta A. Imposing essential boundary conditions in mesh-free methods. *Computer Methods in Applied Mechanics and Engineering* 2004; **193**(12-14):1257–1275.
15. Karlin S, Shapley LS. Geometry of moment spaces. *Memoirs of the American Mathematical Society* 1953; **12**.
16. Tagliani A. Existence and stability of a discrete probability distribution by maximum entropy approach. *Applied Mathematics and Computation* 2000; **110**:105–114.
17. Khinchin AI. *Mathematical foundations of information theory*. Dover: New York, 1957.
18. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal* 1948; **27**(3):379–423.
19. Mead LR, Papanicolaou N. Maximum entropy in the problem of moments. *Journal of Mathematical Physics* 1984; **25**(8):2404–2417.
20. Warren J, Schaefer S, Hirani A, Desbrun M. Barycentric coordinates for convex sets 2005; Submitted.

21. Floater MS, Hormann K, Kós G. A general construction of barycentric coordinates over convex polygons. *Advances in Computational Mathematics* 2005; Accepted.
22. Sukumar N. Construction of polygonal interpolants: A maximum entropy approach. *International Journal for Numerical Methods in Engineering* 2004; **61**(12).
23. Rajan VT. Optimality of the Delaunay triangulation in R^d. *Discrete and Computational Geometry* 1994; **12**(2):189–202.
24. Boyd S, Vandenberghe L. *Convex optimization*. Cambridge University Press: Cambridge, UK, 2004.
25. Attouch H. Viscosity solutions of minimization problems. *SIAM Journal of Optimization* 1996; **6**(3):769–806.
26. Attouch H, Cominetti R. L^p approximation of variational problems in L^1 and L^∞. *Nonlinear analysis* 1999; **36**(3):373–399.
27. Wandzura S, Xiao H. Symmetric quadrature rules on a triangle. *Computers and Mathematics with Applications* 2003; **45**(12):1829–1840.
28. Breitkopf P, Rassineux A, Savignat JM, Villon P. Integration constraint in diffuse element method. *Computer Methods in Applied Mechanics and Engineering* 2004; **193**(12-14):1203–1220.
29. Timoshenko S, Goodier JN. *Theory of Elasticity*. McGraw-Hill: New York, 1951.
30. Organ D, Fleming M, Terry T, Belytschko T. Continuous meshless approximations for nonconvex bodies by diffraction and transparency. *Computational Mechanics* 1996; **18**(3):225–235.
31. Belytschko T, Krongauz Y, Organ D, Fleming M, Krysl P. Meshless methods: An overview and recent developments. *Computer Methods in Applied Mechanics and Engineering* 1996; **139**(1-4):3–47.
32. Krysl P, Belytschko T. Element-free galerkin method: Convergence of the continuous and discontinuous shape functions. *Computer Methods in Applied Mechanics and Engineering* 1997; **148**(3-4):257–277.